UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

G. MALECOT

LES MODÈLES STOCHASTIQUES EN GÉNÉTIQUE DE POPULATION

H. THEIL et L. B. M. MENNES

CONCEPTION STOCHASTIQUE DE CŒFFICIENTS MULTIPLICATEURS DANS L'AJUSTEMENT LINÉAIRE DES SÉRIES TEMPORELLES

M. SKLAR

FONCTIONS DE RÉPARTITION A N DIMENSIONS ET LEURS MARGES

P. THIONET

UN MODÈLE ÉCONOMÉTRIQUE
LA PROJECTION A LONG TERME DE VERDOORN

VOL. VIII - FASCICULE 3 - 1959

PARIS
11, Rue Pierre-Curie

UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

G. MALECOT

LES MODÈLES STOCHASTIQUES EN GÉNÉTIQUE DE POPULATION

H. THEIL et L. B. M. MENNES

CONCEPTION STOCHASTIQUE DE CŒFFICIENTS MULTIPLICATEURS DANS L'AJUSTEMENT LINÉAIRE DES SÉRIES TEMPORELLES

M. SKLAR

FONCTIONS DE RÉPARTITION A N DIMENSIONS ET LEURS MARGES

P. THIONET

UN MODÈLE ÉCONOMÉTRIQUE

LA PROJECTION A LONG TERME DE VERDOORN

VOL. VIII - FASCICULE 3 - 1959

PARIS
11, Rue Pierre-Curie

Toute la correspondance relative aux publications
doit être envoyée à l'adresse
INSTITUT DE STATISTIQUE DE L'UNIVERSITÉ DE PARIS
Institut Henri Poincaré - 11, Rue Pierre Curie - Paris (5e)

Les manuscrits doivent être envoyés à M. Daniel DUGUE
à l'adresse précédente.

LES MODELES STOCHASTIQUES EN GÉNÉTIQUE DE POPULATION

G. MALECOT

INTRODUCTION

La notion de "population infiniment nombreuse"(1) traduisait une approximation qui est parfaitement justifiée dans certaines circonstances : toutes celles où la composition de chaque génération peut être regardée comme parfaitement définie par la composition de la génération précédente. La présente étude va porter au contraire sur le cas d'une "population d'effectif limité", c'est-à-dire d'une population dans laquelle les gamètes qui sont à l'origine de chaque génération F_{n+1}, résulteront d'un nombre limité de "tirages au sort" parmi les gamètes (en nombre illimité) auxquels pourrait théoriquement donner naissance (dans la génération précédente F_n), l'un ou l'autre des systèmes de croisement précédemment définis. Sur les gamètes choisis par ces tirages, ou "gamètes utiles" (mâles ou femelles), la fréquence effective q' du gène a diffèrera en général -du fait du nombre limité de tirages- de la probabilité qu'avait chaque gamète de porter a.

Cette différence traduit l'"effet de hasard" (random drift, genetic drift) dans la reproduction. Cet "effet de hasard" fait de la fréquence q' une "variable aléatoire", mais il nous sera possible de formuler la loi de probabilité de cette variable aléatoire sous les hypothèses suivantes : nous supposerons que les tirages au sort qui fournissent les différents gamètes utiles mâles (ou femelles) sont indépendants et ont chacun une probabilité q_1 de fournir le gène a, q_1 étant une fonction déterminée de la fréquence effective q du gène a parmi les gamètes utiles de la génération précédente, (fréquence qui est la moyenne arithmétique $\frac{q_m + q_f}{2}$ des fréquences des gamètes mâles et femelles).

---

(1) dont l'évolution "déterministe" sous les hypothèses usuelles a été résumée dans mon mémoire de "Population", 1955, n°2 , p. 239.

Nous allons indiquer des conditions suffisantes sous lesquelles ces hypothèses seront vérifiées, conditions qui formeront l'ossature des différents modèles étudiés.

1/ La (faible) sélection et la mutation seront supposées conformes au modèle D de mon mémoire cité. Ce modèle est regardé comme définissant les croisements théoriques, la survie théorique des zygotes et la production théorique des gamètes en nombre illimité, parmi lesquels sont choisis les gamètes utiles à l'origine de F_{n+1}. Ce modèle permettra donc de calculer les probabilités, pour chaque gamète utile, de dériver d'un croisement déterminé, puis d'un zygote déterminé issu de ce croisement, et enfin d'être porteur du gène a ; les probabilités relatives à la production des différents zygotes par chaque croisement et à la production des différents gamètes par ces zygotes, étant parfaitement définies par les lois de Mendel et par la donnée des constantes de sélection, il suffira de connaître, en outre, le système de croisement et la fréquence q de a parmi les gamètes utiles de la génération précédente, pour pouvoir en déduire la probabilité q_1 comme fonction de q et des coefficients de sélection et de croisement) à l'aide de la formule :

$$(D_3)\ q_1 - q = \delta(q) = -uq + v(1-q) + pq\,[ft_1 + (t_0 + w_0 q)(1-f)]$$

Nous nous bornerons, sauf indication contraire, au cas où la même formule est valable pour les gamètes mâles et les gamètes femelles, c'est-à-dire au cas de gènes autosomaux et de coefficient de sélection indépendants du sexe.

2/ Cette probabilité fixe q_1 sera regardée comme attachée au tirage qui détermine chaque gamète utile, quels qu'aient pu être les resultats des tirages ayant préalablement déterminé d'autres gamètes utiles ; autrement dit, les tirages des divers gamètes utiles d'une même génération seront assimilés à des tirages dans une urne à composition constante(1) (où l'on remet les boules tirées). Cette hypothèse paraît fort critiquable : dans le cas théorique, où le nombre des gamètes utiles produit par chaque individu serait invariable -par exemple si ce nombre était égal à 2, ce qui assurerait la constance rigoureuse de l'effectif de la population au cours des générations- il est clair que le tirage préalable d'un certain nombre -2 par exemple- des gamètes utiles, issus d'un même individu, exclurait toute participation de cet individu à la production des autres gamètes utiles (il y aurait analogie avec un tirage d'urnes où l'on ne remet pas les boules tirées). Notre hypo-

---

(1) Il s'agit d'une composition constante pour tous les tirages relatifs à une même génération, car nous verrons que la composition change d'une génération à la suivante.

thèse revient donc à admettre que, aussi longtemps qu'il y a des tirages de gamètes, tous les individus de même génotype de la génération considérée, y participent également, quel que soit le nombre de gamètes utiles qu'ils ont déjà fourni ; cela entraîne que le nombre de gamètes utiles produit par chacun de ces individus, sont des variables aléatoires indépendantes, obéissant à des lois binomiales très voisines pratiquement de lois de Poisson (dont la moyenne k peut naturellement -dans le cas de sélection- dépendre du génotype des individus et du gène porté par les gamètes). Nous avons donc introduit une hypothèse -que l'on peut appeler hypothèse de "fécondité poissonnienne"- qui rendra plus commode nos calculs ultérieurs. Mais nous verrons, dans un autre article, qu'il est facile, moyennant l'introduction dans les formules d'un facteur correctif, de s'affranchir de cette hypothèse restrictive, en englobant l'éventualité où la distribution de gamètes utiles, produits par un même individu, a une variance inférieure ou supérieure à celle (égale à la moyenne k) d'une loi de Poisson (ce qui correspondrait à un schéma d'urnes "hyponormal" ou "hypernormal").

Dans l'hypothèse de "fécondité poissonnienne", il est facile de définir en fonction de q_1 (donc de q) la loi de probabilité de la fréquence q'_m (ou q'_f) du gène a parmi les gamètes utiles mâles (ou femelles) à l'origine de F_{n+1}. Si l'on désigne par N le nombre d'individus comptés à la naissance dans F_{n+1}, N est aussi le nombre de gamètes utiles de chaque sexe à l'origine de F_{n+1}, le nombre Nq'_m (ou Nq'_f) de ces gamètes utiles mâles (ou femelles) porteurs de a obéira à une "loi binomiale", dont les probabilités sont les différents termes du développement de $(p_1 + q_1)^N$. Si l'on regarde N comme une constante, ou bien comme une quantité dont la loi de variation au cours des générations est donnue, (ce qui se justifie toutes les fois que la concurrence écologique impose à l'effectif un niveau sensiblement déterminé, mais exclut, par contre, les cas de "peuplement des places vides" ou de "populations trop peu nombreuses"), la connaissance de la fréquence q fixera complètement la loi de probabilité de la fréquence q'_m (ou q'_f) à la génération suivante, et en particulier, elle fixera sa valeur moyenne -qui sera $q_1 = q + \delta(q)$-, et sa variance, qui sera $p_1 q_1/N$. Puisque nous nous sommes placés dans le cas de gènes autosomaux et de coefficients de sélection indépendants du sexe, q_1 est indépendant du sexe, donc la loi de probabilité, la moyenne, et la variance, seront les mêmes pour les gamètes mâles et les gamètes femelles (les zygotes mâles et femelles qui auront donné respectivement naissance à ces gamètes auront été prélevés parmi les descendants d'un même ensemble de croisements théoriques, et présenteront par suite les mêmes probabilités). Mais l'indépendance qui existe par hypothèse entre tous les gamètes de même sexe, ne peut pas être regardée comme subsistant intégralement entre les gamètes de sexe différents : en effet, à tout gamète utile mâle, correspond un gamète femelle qui s'unit à lui et qui ne peut être regardé

comme choisi indépendamment, car il existe en général entre les 2 gamètes qui s'unissent, une "corrélation gamétique" que nous avons calculée dans 2 cas importants ("Population", 1955, p. 246 et 250) et que nous avons désignée par f. On voit aisément que f est aussi le coefficient de corrélation de q'_m et q'_f. Connaissant f, il est possible de calculer la moyenne et la variance de la fréquence q' du gène a sur l'ensemble des N individus naissants de F_{n+1} ; car q' est évidemment la demi-somme des fréquences de a chez les gamètes utiles mâles et les femeles qui se sont unis pour constituer ces individus : $q' = \frac{q'_m + q'_f}{2}$; q' a donc pour valeur moyenne la valeur moyenne commune de q'_m et q'_f qui est $q_1 = q + \delta(q)$; mais la variance de q' est inférieure à celle $(p_1 q_1/_N)$ de q'_m et q'_f ; elle est en effet égale à $\frac{1}{4}$ [variance (q'_m) + variance (q'_f) + 2 covariance (q'_m, q'_f)] c'est-à-dire à $\frac{1}{4}\left[\frac{2 p_1 q_1}{N} + 2 f \frac{p_1 q_1}{N}\right] = (1 + f)\, p_1 q_1/_{2N}$.

Nous sommes ainsi parvenus à calculer la moyenne et la variance de la loi de probabilité "conditionnelle" de q' (fréquence de a dans F_{n+1}) quand q (fréquence de a dans la génération précédente F_n) est connu.

Si la loi conditionnelle était complètement formulée, elle définirait un schéma de "chaîne de Markoff" (dont elle représenterait la "probabilité de passage") qui relierait les fréquences q, q', ..., $q^{(p)}$, du gène a dans les générations (naissantes) F_n, F_{n+1}, ..., F_{n+p} ... Ce schéma permettrait (en faisant n = 0) d'exprimer, en fonction de la "fréquence initiale" q_0, la loi de probabilité "a priori" de la fréquence dans la $p^{ième}$ génération postérieure. Il résulte de la théorie de chaîne de Markoff que cette loi de probabilité à priori tend, dans le cas "non oscillant", vers une loi limite ("distribution stationnaire") quand le nombre p de générations tend vers l'infini ; cette loi limite existe certainement et est indépendante de la fréquence initiale q_0 dans le "cas régulier", où l'on se trouve toutes les fois que le passage, en un nombre fixe convenable de générations, de toute valeur de la fréquence à une valeur particulière au moins de celle-ci est possible, donc, en particulier, quand il y a sélection et mutation[2]. On peut montrer que, quand il n'y a pas de mutation, il y a encore une loi limite, mais elle

----

(1) Cette variance est donc comprise entre $p_1 q_1/2N$ (panmixie) et $p_1 q_1/N$ (homogamie totale). (Il y a lieu de corriger la formule erronée de G. Malécot (1), p. 46).

(2) Mais non quand il il y a sélection seule, car on peut jamais en l'absence de mutations, passer des valeurs q = 0 ou q = 1 à d'autres valeurs au cours des générations ultérieures.

dépend de la fréquence initiale q_0. Nous allons d'abord, dans un premier modèle A), chercher la moyenne et la variance des lois limites dans le cas où il n'y a pas de sélection.

A - CAS DE GENES NEUTRES SUJETS A MUTATION (et, éventuellement, à immigration en provenance d'un groupe à composition constante).

Puisqu'il n'y a pas hypothèse aucune sélection (génétique ou zygotique), les "constantes de sélection" t_1, t_0, w_0 qui figurent dans la formule (D_3) sont nulles, et (D_3) se réduit à la formule linéaire :

$$(1) \qquad q_1 = q + \delta(q) = q - u\,q + v\,(1 - q)$$

u et v désignant les "taux de mutation" de $\underline{a}$ en $\underline{A}$ et de $\underline{A}$ en $\underline{a}$ (dans le cas où la population étudiée subirait en outre une immigration, à un taux constant m, en provenance d'une autre population où la fréquence de $\underline{a}$ serait une constante q_2, nous avons vu [(op. cit.) p. 252 et 260] qu'il suffit de remplacer u et v par $u' = u + m\,p_2$ et $v' = v + m\,q_2$, quantités que nous appelons les "taux de mutation modifiés" et dont il est bon de remarquer qu'ils peuvent être très supérieurs aux taux de mutation réels u et v).

Du fait de la linéarité, il va nous suffire de la connaissance de la valeur moyenne q_1 et de la variance $(1+f)\,p_1 q_1/2\,N$ de la loi de probabilité conditionnelle de q', pour écrire les formules de récurrence qui unissent soit les valeurs moyennes, soit les variances, des lois de probabilité a priori des fréquences dans les générations successives. Si nous convenons de désigner chaque moyenne a priori par la lettre correspondance surlignée, on obtient immédiatement (en vertu du théorème dit "des moyennes conditionnelles" (1).

$$(2) \qquad \overline{q'} = \overline{q_1} = \overline{q} + \overline{\delta(q)} = \overline{q} - \overline{u\,q} + v(1 - \overline{q})$$

La limite, quand le nombre de générations écoulées augmente indéfiniment, de la moyenne à priori, s'obtient en faisant $\overline{q'} = \overline{q}$ dans (2). Cette limite est donc $C = \frac{v}{u+v}$. Mais la formule (2) montre aussi que cette limite n'est atteinte que très lentement, car on a :

$$\overline{q'} - \frac{v}{u+v} = (1 - u - v)\,(\overline{q} - \frac{v}{u+v})$$

---

(1) La moyenne a priori de la moyenne conditionnelle d'une variable aléatoire est égale à la moyenne de cette aléatoire.

Il faut donc un nombre de générations de l'ordre de grandeur de $\frac{1}{u+v}$ pour diminuer de façon appréciable l'écart entre la moyenne $\overline{q}$ et sa limite C.

Les résultats sont en accord avec ceux que nous avions établis (op. cit.), p. 255, car, dans une population infiniment nombreuse, la fréquence q n'est pas aléatoire et ne diffère donc pas de sa valeur moyenne (notée ci-dessus $\overline{q}$).

Mais dans le modèle stochastique "à effectif limité", que nous étudions présentement, l'ordre de grandeur de l'écart entre la fréquence q et sa moyenne $\overline{q}$ sera fourni par la "variance à priori" $\overline{(q - \overline{q})^2}$ dont la variation au cours des générations peut se déduire de la formule :

$$q' - \overline{q'} = q' - q_1 + q_1 - \overline{q'} = q' - q_1 + (1 - u - v)(q - \overline{q})$$

La moyenne conditionnelle (quand q est connu) du carré du 1er membre, est en effet la somme de la variance conditionnelle $(1+f)\, p_1 q_1 / 2N$ du 2ème membre et du carré $(1 - u - v)^2 (q - \overline{q})^2$ de sa moyenne conditionnelle. En prenant les moyennes à priori, on obtient :

$$\overline{(q' - \overline{q'})^2} = (1 + f)\, \overline{p_1 q_1} / 2N + (1 - u - v)^2\, \overline{(q - \overline{q})^2}$$

On en déduit aisément que la variance a priori tend vers une limite σ^2 dont on obtient la valeur numérique en remplaçant $\overline{(q' - \overline{q'})^2}$ et $\overline{(q - \overline{q})^2}$ par σ^2 et en remarquant que, asymtotiquement :

$$\overline{p_1 q_1} = \overline{q_1} - \overline{q_1^2} = C - C^2 - (1 - u - v)^2 \sigma^2,$$

ce qui fournit :

$$\sigma^2 = \frac{(1 + f)(C - C^2)}{4N(u+v) - 2N(u+v)^2 + (1 + f)(1 - u - v)^2} \tag{3}$$

Cette formule montre que σ qui caractérise la fluctuation de q autour de sa valeur moyenne quand celle-ci est stabilisée et égal à C- est de l'ordre de grandeur de $\frac{1}{\sqrt{4N(u+v) + 1 + f}}$ quand $(u + v)$ est petit.

σ est donc d'autant plus grand que $u + v$, somme des "taux de mutation", est petit par rapport à $\frac{1}{N}$ -inverse de l'"effectif" de la population-.

Un cas particulier important est celui où $u + v$ est pratiquement négligeable, c'est-à-dire où les mutations sont sans effet statistique, même au cours d'un très grand nombre de générations. Dans ce cas la formule (2) (avec $u = v = 0$) donne $\overline{q'} = \overline{q}$, donc la valeur moyenne à priori reste constante et égale à la valeur initiale q_o, on a donc $C = q_o$; et la formule asymptotique (3) se réduit à $\sigma^2 = C - C^2$, ce qui impose à la fréquence q d'être une aléatoire n'ayant comme valeurs possibles que 1 ou 0 avec les probabilités respectives $C = q_o$ et $1 - C = 1 - q_o$. La signification pratique de ce résultat est la suivante : alors que, initialement, les deux gènes a et A coexistaient dans la population (avec des fréquences respectives q_o et $1 - q_o$), l'"effet de hasard" finit à la longue par faire disparaître complètement l'un de ces gènes (qu'aucune mutation, par hypothèse, ne réintroduit de façon durable), de sorte que l'état asymptotique est un état d'"homogénéité génétique" où un seul et même allèle figure dans les loci de tous les individus. Toutefois, la nature de cet allèle unique reste à priori aléatoire : le résultat que nous avons obtenu signifie qu'il y a, à priori, une probabilité q_o pour que ce soit l'allèle a.

Appliquons ceci aux groupes sanguins : les taux de mutation des gènes O, A, B, l'un en l'autre, sont assez faibles pour que la transmission de ces gènes soit à la base des "recherches de paternité" ; or, il est frappant de vérifier que les populations restées pratiquement isolées durant une longue période (les Amérindiens, par exemple) présentent, en ce qui concerne ces gènes, une quasi homogénéité (si cette homogénéité consiste le plus souvent en une présence exclusive du gène 0, cela tient peut-être à ce qu'il est, en moyenne, le plus fréquent dans l'espèce humaine actuelle, et qu'il l'était peut-être aussi pour le groupe humain initial d'où dérivent, par ségrégation, les races actuelles). Le fait que l'hétérogénéité persiste, par contre, sur l'ensemble du globe, doit être attribué à ce que les taux de mutation, quoique négligeables par rapport à l'inverse de l'effectif d'un groupe isolé, ne sont plus négligeables par rapport à l'inverse de l'effectif de la population totale du globe, pour laquelle $N(u+v)$ est alors appréciable, ce qui, d'après la formule (3), entraîne pour σ^2 une valeur nettement inférieure à celle ($C - C^2$) qui implique l'homogénéité (on doit se souvenir que, plus σ^2 est petit, plus la fréquence aléatoire q est, avec grande probabilité, voisine de sa moyenne C -qui est > 0 et < 1- donc plus assuré est le maintien de l'hétérogénéité).

Remarquons que notre raisonnement, quoique ne supposant pas qu'il y ait panmixie -nous avons introduit une "corrélation gamétique" f- à l'inconvénient de n'introduire, comme fréquence q, que la fréquence globale dans toute la population, ou bien la fréquence dans un groupe complètement isolé. Si l'on veut étudier la répartition des fréquences, dans des groupes partiellement isolés qui échangent quelques individus

par migration, il faudra recourir à un autre modèle que nous exposerons ci-dessous § D (pour le généraliser ensuite au paragraphe F). Pour préparer notre étude, que la notion de "coefficient de parenté" rendra assez intuitive, nous allons traiter d'abord un cas simple :

B - CAS PARTICULIER : COEFFICIENT DE PARENTE DANS UN GROUPE PANMICTIQUE ISOLE (sans sélection).

Nous avons défini (op. cit. p. 246) le "coefficient de parenté" de deux individus I_1 et I_2. Le "coefficient de consanguinité" de tout individu se ramène, (à moins d'altération du gène par mutation dans sa transmission de parent à enfant), au coefficient de parenté de son père et de sa mère, ou encore, en vertu de notre hypothèse de panmixie, au coefficient de parenté de deux individus quelconques du groupe ; ces individus I_1 et I_2 sont naturellement supposés pris dans la même génération, que nous noterons F_n (qu'ils soient de sexes différents ou de même sexe n'influe pas sur les probabilités, puisque nous nous bornons à considérer des gènes autosomaux). Mais le coefficient de parenté "à priori", évalué d'après la seule connaissance de la "génération initiale" F_o, ne peut être calculé qu'en regardant I_1 et I_2 comme pris au hasard dans la génération F_n, ou encore en regardant leurs parents respectifs comme pris au hasard dans la génération F_{n-1}.

Les pères, par exemple, de I_1 et I_2 ont donc une certaine probabilité de coincider : cette probabilité est égale à la probabilité pour que deux gamètes mâles, pris au hasard parmi tous les gamètes utiles qui donnent naissance à F_n, soient produits par le même individu ; d'après notre hypothèse de "fécondité poissonnienne", cette probabilité est $\frac{1}{N_m}$, N_m étant le nombre de "pères" distincts figurant dans la génération F_{n-1} ; il en résulte que la probabilité pour que les pères de I_1 et I_2 soient deux individus distincts - J_1 et J_2 - de F_{n-1}, est $1 - \frac{1}{N_m}$. Définissons donc comme <u>coefficient de parenté dans la génération</u> F_n (1), et notons φ_n, la probabilité à priori pour que deux loci homologues quelconques, pris au hasard chez deux individus <u>distincts</u> I_1 et I_2 de F_n, dérivent, <u>sans avoir subi de mutation</u>, d'un même locus d'un ancêtre commun : nous qualifierons dorénavant d'"identiques" de tels loci ; le <u>coefficient de consanguinité</u> d'un individu quelconque de F_{n+1} s'écrit alors $f_{n+1} = (1 - u)^2 \varphi_n$, u étant une moyenne convenablement définie des deux taux de mutation u et v introduits précédemment (2). Et la probabilité pour que I_1 et I_2

----

(1) dans "Population", 1955, p. 246, nous avions défini le coefficient de parenté de deux individus <u>donnés</u>, reliés par des chaînes de parenté <u>connues</u>. Dans le présent article, il s'agit d'individus et de chaînes de parenté aléatoires.

(2) Nous n'insistons pas sur la définition de cette moyenne, car il faut tenir compte de ce que la réalisation successive, sur une même chaîne de parenté, de 2 mutations inverses rétablit le gène primitif ; une étude détaillée sera faite dans un prochain mémoire.

aient reçu deux loci "identiques" de deux pères distincts sera $\frac{1}{4}(1 - \frac{1}{N_m})$ $(1 - u)^2 \varphi_{n-1}$, alors que la probabilité pour qu'ils aient reçu deux loci "identiques" d'un même père sera : $\frac{1}{4}\frac{1}{N_m}(1 - u)^2 \frac{1 + f_{n-1}}{2}$. En tenant compte de ce que deux loci identiques peuvent provenir aussi de deux mères distinctes, ou d'une même mère, ou respectivement d'un père et d'une mère (forcément distincts)[1], on obtient :

$$\varphi_n = \frac{1}{4}\left(\frac{1}{N_m} + \frac{1}{N_f}\right)(1 - u)^2 \frac{1 + f_{n-1}}{2} + \frac{1}{4}\left(4 - \frac{1}{N_m} - \frac{1}{N_f}\right)(1 - u)^2 \varphi_{n-1}$$

Si l'on désigne par N' la moyenne harmonique de 2 N_m et 2 N_m (N' désigne donc l'effectif total en adultes reproducteurs, dans le cas particulier où les effectifs N_m et N_f des deux sexes sont égaux ; dans le cas général, il sera commode d'appeler N' l'effectif génétique) ; et si l'on remplace f_{n-1} par $(1 - u)^2 \varphi_{n-2}$, on a :

(4) $$\varphi_n = \frac{(1 - u)^2}{2 N'} + \left(1 - \frac{1}{N'}\right)(1 - u)^2 \varphi_{n-1} + \frac{(1 - u)^4}{2 N'} \varphi_{n-2}$$

J'ai indiqué ailleurs[2] comment on résout cette "récurrence linéaire". Quand le rang n de la génération augmente indéfiniment, φ_n tend vers une limite φ, qui, étant alors indépendante de n, est fournie immédiatement par l'équation linéaire (4). Bornons-nous à écrire sa "partie principale" (quand u^2 est négligeable) :

(5) $$\varphi \sim \frac{1}{1 + 4 N' u}.$$

Cette formule (5) est, malgré l'apparente différence de points de vue, très proche de la formule (3) :

1/ Si u est négligeable par rapport à $\frac{1}{N'}$, on obtient $\varphi = 1$: certitude (asymptotique) d'identité des gènes, donc certitude d'homogénéité.

---

(1) Hormis les cas d'"hermaphrodisme" ou de "plantes monoïques" où les formules nécessitent des corrections (d'ailleurs aisées).

(2) "Les mathématiques et l'hérédité" (Masson & Cie), p. 35 ; l'équation y est écrite pour f et non pour φ.

2/ dans le cas général, imaginons que l'expérience soit poursuivie indéfiniment, sur un grand nombre de groupes panmictiques isolés de même effectif ; la probabilité à priori, pour qu'un locus quelconque d'une génération future très éloignée porte le gène a, sera la constante que nous avons appelée C, et la probabilité à priori pour que deux loci homologues de 2 individus quelconques de la même génération (très éloignée), dans un même groupe panmictique, soient identiques -ou, encore, au facteur $(1-u)^2$ près, la probabilité pour que les deux loci homologues d'un même individu K d'une génération F_n très éloignée, soient identiques- n'est autre que φ.

Les probabilités à priori pour que l'individu K soit de génotype aa, aA, AA sont donc, (op. cit., p. 246).

$$R = C^2 + \varphi C(1-C), \quad 2Q = 2(1-\varphi)C(1-C), \quad P = (1-C)^2 + \varphi C(1-C)$$

Cependant, dans chacun des groupes, et en raison de la panmixie, les fréquences réelles des 3 génotypes dans la génération F_n seront (si du moins l'effectif n'est pas petit), voisines de q_i^2, $2p_i q_i$, p_i^2, l'indice i et la fréquence réelle q_i de a, variant d'un groupe à l'autre suivant la loi de probabilité asymptotique étudiée en A. (Les fréquences q_i représentent donc les résultats d'épreuves indépendantes effectuées sur la variable aléatoire, dont la loi de probabilité à priori a été étudiée au paragraphe A). Les moyennes arithmétiques sur l'ensemble des groupes, des fréquences réelles q_i et q_i^2 du gène a et du génotype aa, devront être voisines des probabilités C et R ; en reprenant (avec des N_i tous égaux) le calcul fait dans la 1ère partie (p. 253), on constate que la différence $1 = R - C^2$, doit être voisine de la "variance statistique" des fréquences q_i ; on obtient donc, pour valeur approchée de cette variance :

$$1 = R - C^2 = \varphi\, C(1-C) \sim N \frac{C(1-C)}{1+4N'u}$$

Cela concorde bien, (quand u^2 est négligeable), avec la variance de l'aléatoire q, c'est-à-dire avec la formule (3). (Toutefois N' désigne maintenant la demi-moyenne harmonique des effectifs adultes des deux sexes, alors que N désignait l'effectif total à la naissance ; en outre, le remplacement de u + v par u sera expliqué dans une note du paragraphe F).

Il est utile de souligner, une fois de plus, la coexistence d'une panmixie (croisement au hasard) à l'intérieur de chaque groupe et d'un coefficient de consanguinité $\varphi > 0$ dû à ce que les ancêtres d'un individu K, appartenant par hypothèse au même groupe isolé d'effectif limité, ne peuvent être tous distincts ; il est remarquable que ce coef-

ficient caractérise aussi la variance de la fréquence, sur un ensemble nombreux de groupes isolés. Il sera important de reprendre cette analyse, (ci-dessous, en D), dans le cas de groupes non complètement isolés, entre lesquels il y a migration.

C - CAS DE GENES NEUTRES AVEC MUTATIONS ET MIGRATIONS ENTRE DEUX GROUPES PANMICTIQUES G_1 ET G_2.

Nous utiliserons les mêmes notations qu'en B, mais en faisant précéder des indices 1 et 2 les quantités (en particulier les effectifs) correspondant aux groupes G_1 et G_2. Il y a lieu de distinguer les taux de mutation moyens ${}_1u$ et ${}_2u$ dans les deux groupes, car le calcul sera alors applicable au cas d'un groupe exposé à des radiations atomiques, et échangeant des individus avec un autre groupe non exposé (donc à taux de mutation moindres).

Nous supposerons [cf.(op.cit.), p.250] que la probabilité pour qu'un enfant naissant dans le groupe G_i ait un père -ou une mère- né dans le groupe G_k soit l_{ki} (puisqu'il n'y a que 2 groupes $l_{1i} + l_{2i} = 1$).

Nous appellerons φ_{ij} (n) le coefficient de parenté (défini comme en B)de 2 individus distincts I_1 et I_2 de la génération F pris au hasard parmi ceux qui se reproduisent dans les groupes G_i et G_j (1); la probabilité pour que I_1 et I_2 soient nés respectivement dans les groupes G_k et G_h est $l_{ki}\ l_{hj}$; ils ne peuvent avoir de parent commun que si h = k et la probabilité pour qu'ils aient le même père est donc :

$$\frac{1}{{}_1N_m} l_{1i}\ l_{1j} + \frac{1}{{}_2N_m} l_{2i}\ l_{2j}.$$

Il en résulte que la probabilité pour qu'ils aient reçu deux loci "identiques" d'un même père est :

$$\frac{1}{4}(1 - {}_iu)(1 - {}_ju)\left[\frac{1}{{}_1N_m} l_{1i}\ l_{1j}\ \frac{1+\varphi_{11}(n-2)}{2} + \frac{1}{{}_2N_m} l_{2i}\ l_{2j}\ \frac{1+\varphi_{22}(n-2)}{2}\right]$$

alors que la probabilité pour qu'ils aient reçu deux loci "identiques" de deux pères distincts est :

---

(1) Il sera commode de supposer φ_{ij} évalué après toute mutation éventuellement produite dans le germen de I_1 et I_2 pour les environnements des groupes G_i et G_j.
φ_{ij} sera dorénavant appelé le "coefficient de parenté des groupes G_i et G_j", et φ_{ii} le "coefficient de consanguinité du groupe G_i".

$$\frac{1}{4}(1-{}_iu)(1-{}_ju)\Big[\, l_{1i}\, l_{2j}\, \varphi_{12}\,(n-1) + l_{2i}\, l_{1j}\, \varphi_{21}\,(n-1) + \Big(1-\frac{1}{{}_1N_m}\Big) l_{1i}\, l_{1j}\, \varphi_{11}(n-1) +$$

$$+ \Big(1-\frac{1}{{}_2N_m}\Big)\, l_{2i}\, l_{2j}\, \varphi_{22}\,(n-1) \Big]$$

En désignant par ${}_1N'$ et ${}_2N'$ les "effectifs génétiques" des deux groupes on obtient une récurrence qui s'écrit plus brièvement :

$$\varphi_{ij}(n) = 1 - {}_iu)\,(1 - {}_ju)\Big[\sum_{kh} l_{ki}\, l_{hj}\, \varphi_{kh}(n-1) + \frac{1}{{}_1N'}\, l_{1i}\, l_{1j}\Big\{\frac{1+\varphi_{11}(n-2)}{2} -$$

$$-\varphi_{11}\,(n-1)\Big\} + \frac{1}{{}_2N'}\, l_{2i}\, l_{2j}\Big\{\frac{1+\varphi_{22}\,(n-2)}{2} - \varphi_{22}(n-1)\Big\}\Big] .$$

Nous avons montré [1] que ces coefficients de parenté tendent vers des limites φ_{ij}. Ces limites sont données par le système linéaire :

$$(6)\quad \varphi_{ij} = (1-{}_iu)\,(1-{}_ju)\Big[\sum_{kh} l_{ki}\, l_{hj}\, \varphi_{kh} + \frac{1}{{}_1N'}\, l_{1i}\, l_{1j}\, \frac{1-\varphi_{11}}{2} + \frac{1}{{}_2N'}\, l_{2i}\, l_{2j}\, \frac{1-\varphi_{22}}{2}\Big]$$

toujours résoluble, mais dont la solution nécessite trop d'écritures pour que nous la donnions dans le cas général. Nous nous bornerons au cas particulier où le groupe panmictique G_2 est suffisamment vaste pour que son coefficient de consanguinité φ_{22} , et son coefficient de parenté φ_{12} avec G_1, soient négligeables, ainsi d'ailleurs que $\frac{1}{{}_2N'}$.

En désignant [comme en (1), p. 252] par m le coefficient d'immigration de G_2 dans G_1, on a $l_{21} = m$ et $l_{11} = 1 - m$, et la formule (6) donne simplement le coefficient de consanguinité φ_{11} dans le groupe G_1:

$$\varphi_{11} = (1 - {}_1u)^2\,(1 - m)^2\Big[\varphi_{11} + \frac{1}{{}_1N'}\,\frac{1-\varphi_{11}}{2}\Big]$$

Si l'on suppose que m^2 soit, comme ${}_1u^2$, négligeable, cette formule se réduit simplement à :

$$(7)\qquad \varphi_{11} \sim \frac{1}{1 + 4\,{}_1N'\,({}_1u + m)}$$

---

(1) Annales de l'Université de Lyon, A, 1950, p. 48, et 1951, p. 90 et 105 à 107. (Il y a lieu de remplacer $(1-u)^2$ par $(1-{}_iu)(1-{}_ju)$ pour tenir compte de la variation du taux de mutation d'un groupe à l'autre).

qui n'est autre que l'extension de la formule (5) au cas où, à l'effet des mutations, se superpose l'effet d'une immigration constante en provenance d'un groupe très vaste. Pratiquement, il arrivera rarement que le taux de mutation ${}_1u$ dans le groupe G_1, soit aussi du même ordre de grandeur que le "taux d'immigration" m ; c'est donc en général ce dernier qui déterminera la consanguinité dans le groupe G_1 (dans le cas contraire, d'ailleurs, le coefficient de consanguinité serait diminué par les mutations). Si nous appliquons ceci à un groupe G_1 exposé à des radiations atomiques, nous constatons que le danger dû à l'augmentation des taux de mutation, ne réside pas dans la consanguinité du groupe (qui n'est pas augmentée), mais dans l'augmentation au sein du groupe de la fréquence d'équilibre C de chaque gène néfaste : la formule (C 4), donnée dans "Population" (p. 254), ne s'appliquait que si les taux de mutation (désignés alors par u et v) étaient les mêmes dans tous les groupes ; par contre, dans le cas présentement étudié, la fréquence q_2 de a dans le groupe G_2 est différente de la fréquence d'équilibre C dans le groupe G_1 : cette dernière est en général plus grande, car elle est égale [d'après la remarque de "Population" p. 260] a :

$$\frac{{}_1v'}{{}_1u' + {}_1v'} = \frac{{}_1v + mq_2}{{}_1u + {}_1v + m},$$

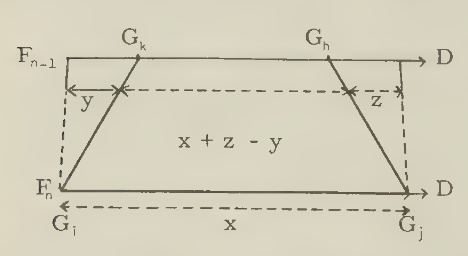
quantité qui est supérieure à q_2 si $(1 - q_2)\,{}_1v > {}_1uq_2$, ce qui a lieu si le gène a est suffisamment rare dans l'ensemble de la population. Si donc la probabilité $R = C^2 + \varphi\, C\,(1 - C)$ des homozygotes aa -seuls porteurs de tares quand celles-ci sont complètement recessives- est augmentée dans le groupe G_1 par les mutations provoquées, cela tient bien à la seule augmentation de C, et non à une augmentation du coefficient de consanguinité φ. Une mesure "eugénique" d'un intérêt évident pour combattre les tares récessives d'origine atomique, consiste en l'augmentation du taux d'immigration m dans le groupe "exposé" G_1 (en particulier par le choix des conjoints à l'extérieur de ce groupe), augmentation qui a pour effet de diminuer à la fois $C = \frac{{}_1v + mq_2}{{}_1u + {}_1v + m}$ et $\varphi = \frac{1}{1 + 4\,{}_1N'\,({}_1u + m)}$. (Il est manifeste que, en contrepartie, cela augmente le danger pour l'ensemble de la population, dans laquelle le gène néfaste se diffuse par sortie hors du groupe G_1 ; de toutes façons le danger total est (à longue échéance d'ailleurs), proportionnel, d'après "Population", p. 261 - au taux de mutation moyen v du gène normal A en le gène "taré" a, donc au nombre total de roentgens absorbé par tous les individus : ce n'est que la répartition du risque entre les différentes familles qui se modifie suivant le degré d'isolement ou de renouvellement).

D - CAS DE GENES NEUTRES AVEC TAUX DE MUTATION CONSTANTS ET MIGRATIONS CONSTANTES ENTRE DE NOMBREUX GROUPES PANMICTIQUES EGAUX.

Désignons les groupes par G_1, G_2, ... G_i, ..., G_j ... et soit N' la valeur (supposée constante) de l'effectif génétique de chaque groupe ; soit u le taux de mutation moyen (supposé maintenant le même pour tous les groupes). Le raisonnement fait en C) peut être repris (avec suppression des indices inutiles) et conduit à une formule qui est une généralisation de (6) (en ce sens que le crochet comporte maintenant autant de termes en $\frac{1}{N'}$ qu'il y a de groupes, chacun de ces termes correspondant à l'éventualité d'un parent commun dans le groupe correspondant). Sous forme condensée, cette formule s'écrit :

$$(8) \qquad \varphi_{ij} = (1 - u)^2 \left[\sum_{kh} l_{kh}\, l_{hj}\, \varphi_{kh} + \frac{1}{N'} \sum_{h} l_{hi}\, l_{hj}\, \frac{1 - \varphi_{hk}}{2} \right]$$

J'ai indiqué (1) la résolution de ce système linéaire, par la méthode des fonctions génératrices, dans le cas où les groupes sont "régulièrement disposés" sur une ligne, ou sur une surface, et où chacun des taux d'immigration l_{hi} ne dépend que de la distance des groupes G_k et G_i. Le principal résultat de cette étude est de montrer que le coefficient de parenté φ_{ij} entre les groupes G_i et G_j est une fonction exponentielle (décroissante) de la distance de ces groupes.



Nous allons étudier en détail le cas particulier où les groupes sont <u>régulièrement disposés sur une droite D, chacun étant repéré par son abscisse</u>. Le coefficient de parenté entre deux groupes G_i et G_j dont la différence des abscisses est x sera noté $\varphi(x)$ [$\varphi(x)$ est, puisque $\varphi(-x) \equiv \varphi(x)$ une fonction "paire" du nombre algébrique x]. Le "taux d'immigration" l_{ki} est supposé être fonction seulement de l'excès (algébrique) y de l'abscisse de G_k sur celle de G_i : nous le noterons l(y) [' taux d'immigration l_{hj} devra alors être noté l(z) si z désigne l'excès algébrique de l'abscisse de G_h sur celle de G_j]. La formule (8) s'écrit alors :

$$(8') \quad \varphi(x) = (1 - u)^2 \left[\sum_{yz} l(y) l(z) \varphi(x + z - y) + \frac{1}{N'} \sum_{z} l(z)\, l(x + z)\, \frac{1 - \varphi(o)}{2} \right]$$

---

(1) Annales de l'Université de Lyon, A, 1950, p. 49 et 11.

Or, cette équation peut être transformée, par une méthode analogue à celle qui a permis à Kolmogoroff de déduire l'équation de la diffusion, à partir des "déplacements au hasard" du mouvement brownien ; si nous regardons la variation de $\varphi(x)$ en fonction de la distance x comme assez "régulière" pour justifier l'existence de dérivées successives de la fonction $\varphi(x)$, et si nous regardons les déplacements s'écartant beaucoup de la moyenne comme assez peu probables pour que, par rapport à la "variance de migration" $\sigma^2 = \sum_y l(y)(y-\bar{y})^2$ [en posant $\bar{y} = \sum_y l(y)\ y$], les moments centrés d'ordre supérieur $\sum_y l(y)(y-\bar{y})^h$ $(h > 2)$ soient négligeables, la formule de Taylor permet de remplacer $\sum_{yz} l(y)\, l(z)\, \varphi(x+z-y)$ par l'approximation :

$$\sum_{yz} l(y)\ l(z) \left[\varphi(x) + (z-y)\,\varphi'(x) + \frac{1}{2}(z-y)^2 \varphi''(x) \right]$$

qui se réduit, compte tenu de ce que $\sum_y l(y) = \sum_z l(z) = 1$ et de ce que z-y peut s'écrire $(z-\bar{y}) - (y-\bar{y})$, à : $\varphi(x) + \sigma^2 \varphi''(x)$

L'équation (8') où les termes en u^2, en $u\sigma^2$ et en $\frac{u}{N'}$ peuvent être négligés, s'écrit alors :

$$(8'') \qquad 2\,u\,\varphi(x) - \sigma^2 \varphi''(x) = \frac{1-\varphi(o)}{2\,N'} \sum_z l(z)\, l(x+z)$$

Le 2ème membre représente -au facteur $\frac{1-\varphi(o)}{8}$ près-la probabilité pour deux individus séparés par la distance x, d'avoir dans la génération précédente un parent commun. Lorsque la distance x est assez grande pour que cette probabilité soit négligeable, l'équation (8'') se réduit à une "équation sans 2ème membre" à coefficients constants, dont la résolution est immédiate : deux solutions particulières pour $\varphi(x)$ sont $e^{\frac{\sqrt{2u}}{\sigma}x}$ et $e^{-\frac{\sqrt{2u}}{\sigma}x}$; le coefficient de parenté cherché, étant borné, ne peut être que proportionnel à la solution d'exposant négatif, on a donc :

$$\varphi(x) = K\, e^{-\sqrt{2u}\,\frac{|x|}{\sigma}}$$

cette formule étant valable lorsque $\frac{x}{\sigma}$ ("distance réduite" évaluée en prenant pour unité l'"écart type de migration" σ) est assez grand. On voit ainsi qu'une diminution importante du coefficient de parenté ne peut se produire que pour une "distance réduite" $\frac{x}{\sigma}$ de l'ordre de $\frac{1}{\sqrt{u}}$,

c'est-à-dire de l'ordre de 1 000, si le taux de mutation u est de l'ordre de 10^{-6}: quand le taux de mutation est faible, la décroissance à distance du coefficient de parenté est très lente. Plus généralement, la formule permet une évaluation approchée du taux de mutation u en remarquant qu'un accroissement de "distance réduite" égal à $\frac{1}{2\sqrt{u}}$ entraîne sensiblement une diminution de moitié du coefficient de parenté.

L'étude de l'équation (8") quand $\frac{x}{\sigma}$ n'est pas grand -et en particulier la recherche du coefficient de consanguinité $\varphi(o)$- est plus délicate (1). Une première méthode consiste à y remplacer la dérivée seconde $\varphi''(x)$ par la "différence seconde" $\frac{1}{h^2}[\varphi(x+h) - 2\varphi(x) + \varphi(x-h)]$ en désignant par h la distance entre deux groupes voisins, et en supposant, pour simplifier, que chaque groupe ne soit alimenté en immigrants, à chaque génération, que par les deux groupes voisins, avec des probabilités respectives égales à m, de sorte que : $l(o) = 1 - 2m$, $l(h) = l(-h) = m$, et $l(nh) = o$ si $|n| > 1$, on a alors $\bar{y} = 0$, $\sigma^2 = 2mh^2$, et $\sum_z l(z)\, l(x+z) = 0$ dès que $|x| > h$, tandis que $\sum_z l^2(z)$ et $\sum_z l(z)\, l(\pm h + z)$, ont pour parties principales (quand m est petit) 1 et 2 m. On peut alors calculer les valeurs $\varphi(nh)$ par récurrence en faisant x = 0, h, 2 h etc. , ce qui donne :

$$\overset{h}{\longleftrightarrow}\overset{h}{\longleftrightarrow} \quad G_{j-1} \quad G_j \quad G_{j+1}$$

$$2u\varphi(o) - 2m[\varphi(h) - 2\varphi(o) + \varphi(-h)] = \frac{1-\varphi(o)}{2N'}$$

$$2u\varphi(h) - 2m[\varphi(2h) - 2\varphi(h) + \varphi(o)] = 2m\frac{1-\varphi(o)}{2N'}$$

$$2u\varphi(2h) - 2m[\varphi(3h) - 2\varphi(2h) + \varphi(h)] = 0 \text{ etc.}$$

La 1ère équation fournit $\varphi(h) = \varphi(-h)$ en fonction de $\varphi(o)$, la 2ème fournit $\varphi(2h)$ et ainsi de suite. Comme les équations, à partir de la 3ème, sont homogènes, $\varphi(nh)$ doit être, pour $n \geqslant 3$, de la forme lk^n, k désignant la racine inférieure à 1 de l'équation $k^2 - 2(1 + \frac{u}{2m})k + 1 = 0$. Ceci exige que $\varphi(2h)$ et $\varphi(h)$ tels que les fournissent les deux premières

---

(1) Une telle étude est indispensable, même si l'on ne veut que déterminer le facteur K que nous avons fait figurer dans l'expression asymptotique de $\varphi(x)$.

équations, soient eux-mêmes dans le rapport k. On en déduit immédiatement l'équation donnant φ (o) :

$$\left[1+\frac{2u}{m}+\frac{u^2}{2m^2}-k\left(1+\frac{u}{2m}\right)\right]\varphi(o)=\left[\frac{1+\frac{u}{2m}}{4N'm}+\frac{1}{2N'}-\frac{k}{8N'm}\right][1-\varphi(o)].$$

Quand m et le rapport $\frac{u}{m}$ sont tous deux petits, φ (o) se réduit sensiblement à $\frac{1}{1+8N'\sqrt{mu}}$; et(pour $n > o$) $\varphi(nh)$ à $\frac{\left(1-\sqrt{\frac{u}{m}}\right)^n}{1+8N'\sqrt{mu}}$ (1);en revenant aux notations $nh = x$ et $2mh^2 = \sigma^2$ et remplaçant $\left(1-\sqrt{\frac{u}{m}}\right)^{|n|}$ par $e^{-\sqrt{\frac{u}{m}}|n|} = e^{-\frac{\sqrt{2u}}{\sigma}|x|}$ on obtient :

(9) $$\varphi(x) = \frac{e^{-\frac{\sqrt{2u}}{\sigma}|x|}}{1+4N'\sqrt{2u}\,\sigma/h}$$

la constante K introduite précédemment coïncide dans ce cas avec φ(o), dont l'expression s'interprète immédiatement comme le coefficient de consanguinité que présenterait le groupe panmictique isolé étudié au paragraphe B formule (5) si son effectif génétique était :

$$N'_1 = N'\sqrt{\frac{2}{u}}\,\frac{\sigma}{h} = 2N'\sqrt{\frac{m}{u}}.$$

N'_1 représente la dimension de l'"isolat" (au sens de S. Wright) qui rendrait compte de la consanguinité φ (o) qui se manifeste à l'intérieur de chaque groupe ; notre hypothèse que $\frac{u}{m}$ est petit entraîne que N'_1 est très supérieur à l'effectif génétique réel N' de chaque groupe. Mais il ne faudrait pas en conclure à une panmixie qui règnerait sur toute l'étendue occupée par $\frac{N'_1}{N'}$ groupes : cette étendue étant égale à $\frac{N'_1}{N'}$ h= $\sqrt{\frac{2}{u}}\,\sigma$, le coefficient de parenté entre deux individus séparés par cette dernière distance est :

$$\varphi\left(\sqrt{\frac{2}{u}}\,\sigma\right) = e^{-2}\,\varphi(o) \# \frac{1}{8}\,\varphi(o).$$

Les résultats numériques que nous venons d'indiquer, dépendent

----

(1) Il y a lieu de remarquer que φ(nh) reste très voisin de φ (o) tant que l'entier n n'est pas grand.

beaucoup moins qu'on ne pourrait le craindre de la loi de migration particulière et discontinue que nous avons adoptée pour résoudre explicitement l'équation (8''); seul importe au premier chef, le nombre de dimensions dans lequel s'effectuent les migrations. Si l'on conserve l'hypothèse de déplacement dans une dimension, sur une droite, par exemple, mais que l'on adopte une loi de migration continue, gaussienne, d'écart-type σ, la somme dans le 2ème membre de (8'') est le "produit de composition" (multiplié par dx) de deux telles lois de gauss, c'est-à-dire $\frac{1}{\sqrt{2\pi}\,\sigma\,\sqrt{2}} e^{-\frac{x^2}{4\sigma^2}} dx$ et N' doit être remplacé par $\delta\,dx$, δ désignant la "densité linéaire" de l'effectif génétique ; l'équation peut alors se résoudre en appliquant aux deux membres une "transformation de Fourier" : la transformée $G(t) = \int_{-\infty}^{+\infty} e^{itx}\,\varphi(x)\,dx$ vérifie alors l'identité $(2u + \sigma^2 t^2)\,G(t) \equiv \frac{1-\varphi(o)}{2\delta} e^{-\sigma^2 t^2}$, et l'inversion de G(t) donne :

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\,G(t)\,dt = \frac{1-\varphi(o)}{4\pi\delta}\int_{-\infty}^{+\infty} \frac{e^{-itx-\sigma^2 t^2}}{2u+\sigma^2 t^2}\,dt$$

et il suffit de se reporter aux tables (1) des transformées de Fourier pour obtenir :

$$(10) \qquad \varphi(x) = \frac{1-\varphi(o)}{4\delta\sigma\sqrt{2u\pi}}\left[e^{2u-\sqrt{2u}\frac{x}{\sigma}} \int_{\sqrt{2u}-\frac{x}{2\sigma}}^{+\infty} e^{-z^2}\,dz + e^{2u+\sqrt{2u}\frac{x}{\sigma}} \int_{\sqrt{2u}+\frac{x}{2\sigma}}^{+\infty} e^{-z^2}\,dz \right].$$

Compte tenu de la petitesse de u, $\varphi(o)$ est alors sensiblement racine de l'équation :

$$\varphi(o) = \frac{1-\varphi(o)}{4\delta\sigma\sqrt{2u\pi}} \times 2\frac{\sqrt{\pi}}{2},$$

ce qui donne :

$$(11) \qquad \varphi(o) = \frac{1}{1 + 4\delta\sigma\sqrt{2u}}$$

alors que, quand $\frac{x}{\sigma}$ est suffisamment grand pour que la dernière intégrale figurant dans (10) soit négligeable, on obtient sensiblement :

---

(1) Erdelyi, Magnus, etc., "tables of integral transforms", Mac Graw Hill.

(12) $$\varphi(x) = \frac{1-\varphi(o)}{4\delta\sigma\sqrt{2u\pi}}\, e^{-\sqrt{2u}\frac{x}{\sigma}} \qquad \sqrt{\pi} = \varphi(o)\, e^{-\sqrt{2u}\frac{x}{\sigma}}$$

Les formules (11) et (12), relatives aux 2 cas extrêmes où $\frac{x}{\sigma}$ est nul ou grand, se trouvent coïncider exactement avec la formule (9) établie dans le cas d'une loi de dispersion discontinue, à condition de substituer dans celle-ci, à l'"effectif moyen par unité de longueur" N'/h, la densité linéaire δ.

L'examen que j'ai fait, dans un autre mémoire (1) du cas de la migration "homogène" suivant deux dimensions (migration "en surface"), fait apparaître que $\varphi(x)$ conserve la même allure asymptotique quand $\frac{x}{\sigma}$ est grand, mais que, par contre, sa valeur numérique quand $x = 0$ est assez différente de la valeur (11) obtenue dans le cas d'une seule dimension (migration "linéaire"). Cela tient à ce que la probabilité, pour deux individus voisins, d'avoir un ancêtre commun (plus ou moins éloigné dans le temps), dépend essentiellement, pour une même densité, du nombre de dimensions dans lequel ils se déplacent. L'effet cumulatif de différentiation (et par suite de diminution du coefficient de parenté) que produisent les mutations survenues dans chaque descendance de cet ancêtre commun dépend donc, lui aussi, du nombre de dimensions. Pour le cas d'une loi de migration continue gaussienne suivant deux dimensions, le coefficient de consanguinité $\varphi(o)$ dépend du taux de mutation u, non plus par sa racine carrée, mais par son logarithme, suivant la formule :

$$\varphi(o) = \frac{1}{1 + 8\pi\delta\sigma^2\,(-\,1/\log 2u)} \qquad (2)$$

(σ désigne l'écart-type de la composante parallèlement à une direction fixe du déplacement ; le "déplacement quadratique moyen" serait donc $\sigma\sqrt{2}$).

Alors que le coefficient de parenté à grande distance ($\frac{x}{\sigma}$ grand) est régi par la formule :

$$\varphi(x) = \frac{K}{\sqrt{x}}\, e^{-\sqrt{2u}\frac{x}{\sigma}}$$

---

(1) Annales de l'Université de Lyon, Math. 1950, p. 52 et suiv.

(2) L'augmentation du nombre de dimensions entraîne donc, pour une valeur donnée, de $2\delta\sigma$ ou de $2\pi\delta\sigma^2$ (nombre d'individus dans un cercle ou un segment ayant pour rayon le déplacement quadratique moyen), une diminution de $\varphi(o)$.

Par contre, il semble que la migration discontinue en surface, soit plus proche de la migration suivant une dimension que de la migration continue suivant deux dimensions (1).

Les formules que nous avons obtenues pour $\varphi(o)$ et $\varphi(x)$ sont importantes, du fait qu'elles permettent de prévoir aisément les variances et covariances "intergroupes", relatives aux fréquences q du gène a dans les différents groupes (ce que nous avions déjà constaté, en B, pour la variance sur un grand nombre de groupes complètement isolés ; mais la covariance entre deux groupes distincts était alors nulle). En effet, l'ensemble des loci homologues des individus constituant le groupe G_i, peut être regardé comme résultant d'un ensemble de tirages au sort non indépendants, deux tirages quelconques ayant toujours la probabilité $\varphi(o)$ de fournir deux loci identiques [et la probabilité $1 - \varphi(o)$ de fournir deux loci sans ancêtre commun ou altérés par mutations(2)]. Chaque locus ayant asymptotiquement [comme en B] la probabilité constante C de porter le gène a, la fréquence q_i de ce gène dans le groupe G_i aura la valeur moyenne C, et sa variance sera celle de la moyenne arithmétique d'un très grand nombre (3) de variables aléatoires, ayant chacune la probabilité C de prendre la valeur 1, et ayant, deux à deux, un coefficient de corrélation moyen égal à $\varphi(o)$; cette variance est $C(1 - C)\varphi(o)$ (résultat à rapprocher de celui obtenu en B, et de l'interprétation statistique donnée dans "Population", 1955, p. 253, où C est remplacé par la moyenne générale des fréquences). Pour la même raison, la covariance entre les fréquences q_i et q_j dans 2 groupes distincts G_i et G_j, séparés par une distance x, est $C(1 - C)\varphi(x)$; d'où résulte que le coefficient de corrélation entre q_i et q_j est $\frac{\varphi(x)}{\varphi(o)}$. (Nous avions déjà déterminé la covariance en 1948, dans le cas continu à 2 dimensions, dans notre livre" Les mathématiques de l'hérédité, p. 60, note a) ; la décroissance exponentielle (à distance suffisante) de ce coefficient de corrélation, permet donc (comme Lamotte l'a souligné dans sa thèse sur les populations naturelles de Cepaea Nemoralis) une détermination expérimentale de $\sqrt{\frac{2u}{\sigma^2}}$ ou de $\sqrt{\frac{u}{m}}$ (voir § F).

Dans le cas d'un caractère "biométrique", conditionné par plu-

----

(1) Voir G. Malécot, op. cit. et M. Lamotte Bull. Biol. France et Belgique, suppl. 35.

(2) loci que nous pouvons alors regarder comme indépendants, ce qui entraîne que le coefficient de corrélation moyen entre deux loci se réduit à $\varphi(o)$.

(3) à condition que soient suffisamment nombreux, non seulement l'effectif de chaque groupe G_i, mais encore les individus qui y sont prélevés pour estimer la fréquence q_i ; dans le cas contraire, voir op. cit, 1950, p. 57.

sieurs couples de gènes allèles à effets additifs, (sans dominance ni epistasis), et si ces gènes sont supposés tous neutres et de taux de mutation sensiblement égaux, le coefficient de corrélation entre les valeurs de ce caractère, chez deux individus appartenant à des groupes séparés par une distance x, a une expression un peu différente :

$$\frac{2\ \varphi(x)}{1 + \varphi(o)} \text{ (1)} .$$

Nous pouvons conclure que la parenté qui existe entre les individus constituant des groupes d'effectifs limités et géographiquement rapprochés, entraîne, entre les fréquences dans ces groupes, aussi bien qu'entre les caractères biométriques de leurs individus, l'existence de corrélations positives même dans l'état "stationnaire" asymptotique (à la différence de ce que nous avions établi dans "Population" (p. 254 et 255), pour une population infiniment nombreuse où les fréquences finiraient à la longue par s'uniformiser, par un effet de moyenne, sur toute l'étendue occupée par l'espèce). La "consanguinité de position" (2) présente donc, dans le cas présentement étudié, un caractère permanent ; elle doit, par suite, pouvoir être constatée expérimentalement dans toute population occupant une aire suffisamment étendue (que cette population soit continue, ou fragmentée en groupes panmictiques reliés par des courants de migration).

Mais l'inconvénient du présent modèle est de ne pas tenir compte d'un autre facteur de différentiation entre groupes, facteur souvent plus important que les mutations : la sélection par les conditions locales. Cette nécessaire généralisation sera l'objet du paragraphe F, mais il nous faut, au préalable, donner quelques indications (paragraphe E) sur l'effet de la sélection dans un groupe isolé.

E - SELECTION ET MUTATIONS DANS UN GROUPE ISOLE.

Reprenons les notations du paragraphe A, en désignant par q_1 et par $\omega(q_1) = \frac{(1+f)\ p_1 q_1}{2N}$ la moyenne et la variance de la loi de probabilité conditionnelle de la fréquence q' du gène a dans F_{n+1}, quand sa fréquence q dans F_n est connue. q_1 s'exprime sous la forme $q + \delta(q)$, la quantité $\delta(q)$ traduisant maintenant l'action combinée de la sélection et des mutations et étant fournie par la formule (D_3) de la p. 2.

----

(1) Cela tient à ce que la variance des individus n'est plus, comme celle des fréquences des groupes, proportionnelle à $\varphi(o)$, mais à $1 + \varphi(o)$ (le tirage d'un individu équivaut à celui de deux gènes allèles seulement). Pour les détails, voir op. cit. 1950, p. 59.

(2) M. Lamotte, op. cit.

Nous désignerons par $\Phi_n(q)\,dq$ la probabilité a priori (probabilité qui ne dépend que de la "fréquence initiale" q_o dans la génération F_o), pour que la fréquence dans F_n appartienne à l'intervale infiniment petit $q \ldots q + dq$. $\Phi_{n+1}(q')\,dq'$ désignera de même la "probabilité élémentaire a priori" de la fréquence dans F_{n+1}. Si $\theta(q'q')\,dq'$ représente la probabilité de passage de la valeur q a une valeur de l'intervalle $q' \ldots q' + dq'$, les règles des probabilités totales et composées donnent (après simplification par dq') :

$$\Phi_{n+1}(q') = \int_o^1 \theta(q.\,q')\,\Phi_n(q)\,dq \tag{13}$$

Si nous désignons par $F(s,\,n) = \int_o^1 e^{sq}\,\Phi_n(q)\,dq$ la fonction caractéristique (a priori) de la distribution de probabilité de q, il suffit de multiplier les deux membres de (13) par $e^{sq'}\,dq'$ et d'intégrer pour obtenir (après interversion d'intégration) :

$$F(s,\,n+1) = \int_o^1 \left[\int_o^1 e^{s(q'-q)}\,\theta(q,q')\,dq'\right] e^{sq}\,\Phi_n(q)\,dq.$$

L'intégrale entre crochets pourrait être appelée, puisque la valeur de q y est fixée, la "fonction caractéristique conditionnelle" de l'accroissement aléatoire $q' - q$; elle fournit, par développement, suivant les puissances de s, les "moments conditionnels" de cette aléatoire ; le premier de ces moments n'est autre que $q_1 - q$, c'est-à-dire $\delta(q)$; le deuxième moment n'est autre que $\omega(q_1) + \delta^2(q)$ quantité que nous réduirons à $\omega(q) = \frac{(1+f)\,q\,(1-q)}{2\,N}$ en supposant, avec Sewall Wright, que $\frac{1}{N}$, u, v, t_1, t_o, w_o soient des infiniment petits du même ordre de grandeur, et que leurs carrés et leurs produits mutuels soient négligeables (c'est une extension de l'hypothèse de "faible sélection" faite dans "Population", p. 257) ; ces hypothèses ont l'intérêt de rendre négligeables les moments d'ordre supérieur à 2, donc de réduire pratiquement la "fonction caractéristique conditionnée" à $1 + s\,\delta(q) + \frac{s^2}{2}\,\omega(q)$, ce qui donne alors :

$$F(s,\,n+1) = \int_o^1 \left[1 + s\,\delta(q) + \frac{s^2}{2}\,\omega(q)\right] e^{sq}\,\Phi_n(q)\,dq \tag{14}$$

Le 2ème membre -où $\delta(q)$ et $\omega(q)$ sont des polynomes en q- peut s'exprimer en fonction des dérivées partielles :

$$\frac{\partial^p F(s, n)}{\partial s^p} = \int_0^1 q\, e^{sq}\, \Phi_n(q)\, dq$$ (1)

mais on peut aussi l'écrire, grâce à une intégration par parties :

$$F(s, n) + s \int_0^1 1^{sq} \left\{ \delta(q)\, \Phi_n(q) - \frac{1}{2}\frac{\partial}{\partial q}\left[\omega(q)\, \Phi_n(q)\right]\right\} dq$$

(à condition que la fonction entre crochets soit nulle aux bornes 0 et 1).

Quand la distribution asymptotique stationnaire, indiquée dans l'introduction est atteinte, F (s, n + 1) doit être identique à F (s, n), ce qui nécessite la nullité de l'accolade ; la distribution stationnaire de probabilité de q -distribution dont nous désignerons la densité par $\Phi(q)$- vérifie donc l'équation différentielle :

$$\frac{1}{2}\frac{d}{dq}[\omega(q)\, \Phi(q)] - \delta(q)\, \Phi(q) = 0$$

que l'on peut intégrer par :

(15) $$\omega(q)\, \Phi(q) = K \exp.\left[2 \int \frac{\delta(q)}{\omega(q)}\, dq\right]$$

la constante K étant déterminé par la condition $\int_0^1 \Phi(q)\, dq = 1$.

En remplaçant $\delta(q)$ et $\omega(q)$ par leurs expressions, on trouve :

(16) $$\Phi(q) = K_1\, q^{\frac{4Nv}{1+f}-1}(1-q)^{\frac{4Nu}{1+f}-1} \exp\left\{\frac{2N}{1+f} w_0(1-f)\, q^2 + \frac{4N}{1+f}\left[t_0(1-f)+ft_1\right] q\right\},$$

(à condition que $\omega(q)\Phi(q)$ soit nulle aux bornes de l'intervalle 0... 1, et que $\Phi(q)$ soit intégrable dans cet intervalle, ce qui a lieu quand u et v sont tous deux > 0, c'est-à-dire quand chaque gène est passible de mutations ; dans le cas contraire où il n'y a pas de mutations, la distribution asymptotique ne peut admettre de densité de probabilité, car il y a toujours extension à toute la population de l'un ou l'autre des gènes : q ne peut prendre asymptotiquement que les valeurs 0 ou 1, avec des

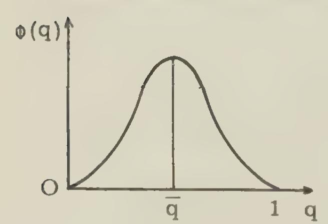
---

(1) Ce qui est à l'origine de la méthode que j'ai exposée dans les "publications de l'Institut de Statistique de Paris", 1952, vol. 1, fasc. 3, méthode qui permet, dans le cas particulier de gènes neutres, le calcul successif de tous les moments (le calcul des moments d'ordre 1 et 2 à déjà été fait ci-dessus, paragraphe A).

probabilités qui dépendent cette fois, de la fréquence initiale q_o [1].

La formule (16), qui fut découverte (dans le cas où f = 0) par Sewall Wright [2] est bien connue des biologistes, car elle fournit à la fois : la probabilité $\phi(q)\,dq$ pour qu'un gène donné a ait, dans un groupe isolé comprenant N individus et au bout d'un nombre suffisamment grand de générations, une fréquence comprise entre q et q + dq ; les proportions asymptotiques des différentes valeurs simultanées de la fréquence q sur un grand nombre de groupes isolés comprenant chacun N individus (en supposant fixes les coefficients de mutation et de sélection) ; et même (d'après le théorème "ergodique") la distribution des valeurs de la fréquence q sur une très longue période, dans un seul groupe isolé. Des tentatives de vérification expérimentale de cette formule ont été faites (voir Lamotte, Bull. Biol. Fr. et Belg., suppl. 35, 1951 ; Wright et Kerr, Evolution, 8, p. 225). Les critiques de la formule (16), faites par l'Ecole de R. A. Fisher, contestent la validité de l'hypothèse de constance des coefficients de sélection dans le temps et dans l'espace (Fisher et Ford, Heredity, 1, p. 143 et 4, p. 117 ; Sheppard, "Evolution as a process", p. 201). Nous indiquerons ci-après le modèle de Kimura comportant une variation aléatoire dans le temps des coefficients de sélection dans un groupe isolé, puis, ci-dessous en F), notre schéma comportant une variation aléatoire de groupe à groupe.

Mais nous rappellerons d'abord les conséquences classiques de la formule (16) ; sans discuter tous les cas possibles, nous nous bornerons à indiquer les 2 cas les plus importants :



1er cas : taux de mutation u et v assez grands vis-à-vis de $\frac{1}{N}$ pour que $\frac{4\,Nu}{1+f}$ et $\frac{4\,Nv}{1+f}$ soient nettement plus grands que 1.

---

(1) probabilités, dont j'ai établi les valeurs :

$$1 - \bar{P} \text{ et } \bar{P} = \frac{1 - \exp\left[-4N\sigma q_o/(1+f)\right]}{1 - \exp\left[-4N\sigma/(1+f)\right]}$$

en posant $\sigma = ft_1 + (1-f)t_o$ (dans le cas particulier où $w_o = 0$), dans le mémoire ci-dessus indiqué.

(2) pour un aperçu des toutes dernières extensions de cette formule, consulter S. Wright, Cold Spring Harbor Symposia on quantitative biology, vol. XX, p. 16.

La courbe représentative de la densité Φ (q) est alors une "courbe en cloche" et la distribution est <u>concentrée</u> dans le voisinage de la valeur moyenne $\bar{q}$ qui est, par extension de la formule (2) du paragraphe A, racine de l'équation $\delta(\bar{q}) = 0$. $\bar{q}$ ne serait autre que la valeur d'équilibre [1] dans une population infiniment nombreuse, valeur étudiée dans un cas particulier, avec la notation r, dans "Population", p.260. La "fluctuation" (dans un grand nombre de groupes isolés, ou au cours du temps dans un même groupe) des valeurs de q autour de $\bar{q}$ s'explique par les déviations que le tirage au sort des gamètes ("random drift"). en se répétant à chaque génération, provoque par rapport à la position d'équilibre $\bar{q}$ que tendent à imposer conjointement les "forces de rappel" dues aux mutations et à la sélection, (rappelons que dans le paragraphe A), en l'absence de sélection, cette dispersion a pu être caractérisée aisément par la variance de q). Quand il s'agit de fluctuations suffisamment petites autour de $\bar{q}$, il est possible, en 1ère approximation, de remplacer la "force de rappel" $q_1 - q = \delta(q)$ définie par la formule (D_3) par une force de rappel "linéaire" [2] de la forme $q - q_1 = -h(q-\bar{q})$, la constante h étant (d'après la formule de Taylor) donnée par :

$$(17)\quad h = -\frac{d\,\delta(\bar{q})}{d\bar{q}} = u + v - [f t_1 + t_0(1-f)]\,(1-2q) - w_0(1-f)(2\bar{q} - 3\bar{q}^2).$$

Cette remarque permet, entre autres, un calcul approché de la variance σ^2 de $\bar{q}$: il suffit d'utiliser le calcul qui avait été fait au paragraphe A dans le cas d'une force de rappel linéaire (cas particulier "sans sélection" : $h = u + v$) : σ^2 est donné par la formule (3), à condition d'y remplacer $u + v$ par h et C par $\bar{q}$ (dans le paragraphe F, cette méthode approchée sera étendue au cas de groupes échangeant des individus par migration).

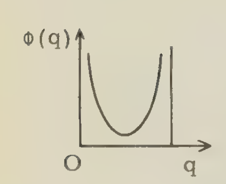
<u>2ème cas</u> : taux de mutation u et v assez petits par rapport à $\frac{1}{N}$ pour que $\frac{4Nu}{1+f}$ et $\frac{4Nv}{1+f}$ soient plus petits que <u>1</u>.

Dans ce cas, la formule (16) montre que $\Phi(q)$ devient infini pour $q = 0$ et $q = 1$: la constante K_1 est d'ailleurs d'autant plus petite que u et v sont plus petits. Ce sont les fréquences proches de $q = 0$ ou de $q = 1$ qui ont la plus forte probabilité. La petitesse des taux de mutation dans le cas que nous considérons, a donc pour conséquence que la plupart des gènes sont près de la fixation ou de la disparition (éventua-

---

(1) Nous laissons de côté les cas (rarement rencontrés quend les coefficients de sélection sont fixes) où cette valeur d'équilibre n'est pas unique.

(2) ce remplacement est d'ailleurs valable en toute rigueur dans le cas de gènes "neutres", pour lesquels la force de rappel n'est due qu'aux mutations.



lités qui se produisent d'ailleurs de temps à autre); si la fixation ou la disparition ne sont pas permanentes, cala tient aux quelques mutations qui surviennent de temps en temps dans un sens ou dans l'autre, ou bien à des immigrations sporadiques s'il s'agit d'un groupe non complètement isolé (par exemple, s'il s'agit des peuplades primitives dont nous avons signalé, au paragraphe A), la quasi-homogénéité concernant le groupe sanguin).

Modèle de Kimura.

La sélection est seulement gamétique, t_0 et t_1 se réduisant à s-1 et ω_0 à 0 ; mais le coefficient de sélection gamétique s, au lieu d'être constants, est aléatoire (de moyenne $\overline{s}$, de variance V_s)[1].

Si l'on note encore $q_1 = q + \delta(q)$ la moyenne conditionnelle de q' (relatif à F_{n+1}) quand q (relatif à F_n) est connu, la formule D_3 donne, en y remplaçant le terme de sélection (aléatoire) par sa valeur moyenne :

$$q_1 - q = \delta(q) = -uq + v(1-q) + (\overline{s}-1)\,q(1-q)$$

Mais la variance conditionnelle de q' est maintenant la somme de la variance $\frac{(1+f)\,p_1 q_1}{2N}$ due au tirage au sort des gamètes, et de la variance $V_s\ q^2(1-q)^2$ du terme aléatoire qui figure maintenant dans (D_3). En désignant par $\omega(q)$ cette somme, la formule (15) reste valable pour déterminer la densité $\Phi(q)$ de la distribution stationnaire. Kimura a montré (quand $u = v$) que la modification apportée à la forme de distribution par le caractère aléatoire de s, n'est importante que si V_s dépasse l'ordre de grandeur de u ou de $\frac{1}{N}$: une distribution "concentrée" peut être remplacée par une distribution "bimodale"[2] dans le cas où $V_s > 4u - (1/N)$; la fluctuation du coefficient de sélection a donc (qualitativement) le même effet que la petitesse des taux de mutation. Pour les détails, nous renvoyons à son intéressant mémoire (qui indique également des extensions).

---

(1) le mémoire de Kimura, dans "Cold Spring Harbor Symposium", vol. XX, p. 39, désigne par s ce que nous notons s-1.

(2) la concentration de la densité vers les 2 extrêmes n'est toutefois pas exactement la même que dans une "distribution en U", car la densité tombe à zéro pour les deux valeurs extrêmes : q = 0 et q = 1.

F - CAS DE MIGRATIONS CONSTANTES ENTRE DE NOMBREUX GROUPES A TAUX DE MUTATION ET DE SELECTION VARIABLES.

Nous désignerons -comme dans "Population", 1955, p. 250- par l_{ki} le pourcentage, à chaque génération, d'individus qui se reproduisent dans le groupe G_i tout en étant nés eux-mêmes dans le groupe G_k.

Nous désignerons par q_k la fréquence du gène a parmi les gamètes utiles qui donnent naissance à la génération F_n du groupe G_k, et par q'_i la fréquence de a parmi les gamètes utiles à l'origine de la génération suivante F_{n+1} du groupe G_i. S'il n'y avait ni mutation ni sélection, la moyenne conditionnelle de q'_i, quand les q_k sont connus, serait simplement $q_{im} = \sum_k l_{ki}\, q_k$; mais cette valeur q_{im} doit en réalité être modifiée par l'addition d'un terme $\delta_i(q_{im})$ qui traduit l'action combinée des mutations et de la sélection.

Nous admettrons, d'une part, que ce terme ne dépend que des conditions régnant dans le groupe G_i, d'autre part, qu'il est linéaire (1) en q_{im}, donc de la forme :

$$\delta_i(q_{im}) = -h_i(q_{im} - C_i) = -h_i\left(\sum_k l_{ki}\, q_k - C_i\right) \tag{18}$$

(h_i se réduisant à $u_i + v_i$ et C_i à $\dfrac{v_i}{u_i + v_i}$ dans le cas où il n'y a pas de sélection et où les taux de mutation dans le groupe G_i sont des taux fixes u_i et v_i).

q'_i admet alors (quand les q_k sont connus), la moyenne conditionnelle :

$$q_{ic} = q_{im} + \delta_i(q_{im}) = h_i\, C_i + (1 - h_i) \sum l_{ki}\, q_k \tag{19}$$

La fluctuation de q'_i autour de q_{ic} sera exprimée par la variance conditionnelle qui, si les taux de mutation et de sélection étaient fixes, s'exprimerait -sous les hypothèses indiquées dans l'introduction- par $\dfrac{(1 + f_i)\, q_{ic}(1 - q_{ic})}{2N_i}$, N_i désigant le nombre d'individus (comptés à la naissance de la génération F_{n+1}) dans le groupe G_i, et f_i la corrélation gamétique dans ce groupe ; pour simplifier, nous réduirons cette variance à $\dfrac{(1 + f_i)\, q_{im}(1 - q_{im})}{2N_i}$, ce qui est légitime si $q_{ic} - q_{im}$ est suffi-

----

(1) Nous avons indiqué dans le paragraphe E (1er cas) dans quel cas le remplacement du terme de sélection par un terme linéaire est légitime.

samment petit, donc si les taux de mutation et de sélection sont suffisamment petits (1).

Mais nous allons nous placer, plus généralement, dans "l'hypothèse de Kimura", en admettant que les taux de mutation et de sélection, au lieu d'être fixes dans chaque groupe, sont aléatoires : les quantités h_i et C_i que nous avons introduites dans la formule (18), correspondront alors seulement aux valeurs moyennes de ces taux dans le groupe G_i ; la variance conditionnelle est la somme du terme indiqué ci-dessus et d'un terme supplémentaire, traduisant la fluctuation des taux réels à chaque génération autour de leurs valeurs moyennes. Cette fluctuation peut être exprimée à l'aide de différentes hypothèses. L'une d'elles -assez justifiée- consiste à regarder l'expression réelle de $\delta_i\,(q_{im})$ à chaque génération comme résultant de deux forces de rappel antagonistes, respectivement proportionnelles à $1 - q_{im}$ et q_{im}, les coefficients de proportionalité étant des variables aléatoires indépendantes v_i et $-u_i$ de valeurs moyennes $\overline{v}_i$ et $-\overline{u}_i$ et de variances V_v et V_u. La moyenne conditionnelle $\delta_i\,(q_{im}) = q_{ic} - q_{im}$ de l'aléatoire $q'_i - q_{im}$ est alors $\overline{v}_i\,(1 - q_{im}) - \overline{u}_i\, q_{im}$, elle est bien toujours donnée par la formule (18) à condition de poser $h_i = \overline{u}_i + \overline{v}_i$ et $C_i = \dfrac{\overline{v}_i}{\overline{u}_i + \overline{v}_i}$. Mais la variance conditionnelle sera maintenant, suivant un raisonnement analogue à celui de Kimura :

(20) $$\omega_i\,(q_{im}) = \frac{(1 + f_i)\, q_{im}\,(1 - q_{im})}{2\,N_i} + V_v\,(1 - q_{im})^2 + V_u\, q_{im}^2$$

La formule (19) permet -en prenant les moyennes a priori des deux membres - d'obtenir la relation entre les moyennes a priori de deux générations consécutives. Elle définit en particulier les valeurs asymptotiques stationnaires $\overline{q_k}$ de ces moyennes a priori par les équations linéaires :

(21) $$\overline{q_i} = h_i C_i + (1 - h_i) \underset{k}{\ } l_{ki}\, \overline{q}_k$$

Le calcul des variances et covariances a priori se fait ensuite, -pour l'état stationnaire- en déduisant de (19) et (21), les formules :

$$q'_i - \overline{q}_i = q'_i - q_{ic} + q_{ic} - \overline{q}_i = q'_i - q_{ic} + (1 - h_i) \sum_k l_{ki}\,(q_k - \overline{q}_k)$$

$$q'_j - \overline{q}_j = q'_j - q'_{je} + (1 - h_j) \sum_n l_{nj}\,(q_n - \overline{q}_n)$$

Il suffit de multiplier membre à membre, de prendre les moyen-

---

(1) Cette approximation n'est pas indispensable, mais elle allège les écritures.

nes conditionnelles [celles de $q'_i - q_{ic}$, de $q'_j - q_{je}$, et du produit de ces quantités par $j \neq i$ sont nulles ; celle de $(q'_i - q_{ic})^2$ est la quantité ω_i définie par (20)], et de désigner par v_{ij} les valeurs stationnaires des moyennes a priori des produits $(q'_i - \overline{q}_i)(q'_j - \overline{q}_j)$ -moyennes qui sont les covariances- pour avoir :

$$(22) \qquad v_{ij} = \sum_{kn} (1 - h_i)(1 - h_j)\, l_{ki}\, l_{nj}\, v_{kn} + \delta_{ij}\, \overline{\omega_i (q_{im})},$$

$$\text{avec : } \overline{\omega_i (q_{im})} = \left[\frac{(1 + f_i)(\overline{q_{im}} - \overline{q^2_{im}})}{2 N_i} + V_u\, \overline{q^2_{im}} + V_v\, \overline{(1 - q_{im})^2} \right]$$

et δ_{ij} étant nul quand $j \neq i$ et égal à 1 dans le cas où $j = i$ (c'est-à-dire dans le cas de la variance v_{ii}).

L'équation (22) -dont l'analogie de forme avec les équations (6) et (8) s'explique par la relation, soulignée à la fin du paragraphe B, entre variances et coefficients de parenté -peut être explicitement résolue dans les mêmes cas particuliers que l'équation (8), et met alors en évidence la décroissance de la covariance entre deux groupes G_i et G_j en fonction de leur distance. Mais, pour obtenir une formule statistiquement vérifiable, il faut tenir compte de ce que les moyennes a priori $\overline{q}_i$, différentes d'un groupe à l'autre, ne pourront qu'exceptionnellement être déterminées individuellement (les fluctuations de la fréquence q'_i dans le groupe G_i pouvant rarement être observées sur un nombre de générations suffisant pour estimer $\overline{q}_i$). Il faut donc considérer, non pas les covariances v_{ij} par rapport aux moyennes vraies, mais les "comoments" $V_{ij} = M[(q'_i - \overline{q}_o)(q'_j - \overline{q}_o)] = v_{ij} + (\overline{q}_i - \overline{q}_o)(\overline{q}_j - \overline{q}_o)$ par rapport à une "moyenne conventionnelle" que nous notons q_o. Et ces comoments ne pourront être estimés (hormis le cas rare ou il serait possible d'étudier chaque couple de groupes pendant un nombre de générations suffisamment élevé) qu'à l'aide de moyennes arithmétiques de produits $(q'_i - \overline{q}_o)(q'_j - \overline{q}_o)$ formés à partir des fréquences q'_i et q'_j observées simultanément sur un grand nombre de couples de groupes G_i et G_j situés à une distance donnée $\pm$ x l'un de l'autre. Nous désignerons par $\psi(x)$ une telle moyenne arithmétique, par $\varphi(x)$ celle des covariances v_{ij} correspondantes, et par $\mu(x)$ celle des produits $(\overline{q}_i - \overline{q}_o)(\overline{q}_j - \overline{q}_o)$ correspondants. On a évidemment :

$$(23) \qquad \psi(x) = \varphi(x) + \mu(x).$$

Nous nous placerons dans le cas où les groupes sont "régulièrement disposés" suivant le schéma exposé au paragraphe D dans le cas discontinu [1] les immigrants dans G_j à chaque génération ne provenant

---

(1) On obtiendrait le même résultat dans le cas continu, à condition de remplacer 2 m par σ^2 et de substituer au terme en δ_{ij} qui figure dans le 2ème membre de (22), une "fonction de Dirac".

-dans un pourcentage m- que de chacun des deux groupes voisins (dont la distance à G_j sera maintenant, pour simplifier, prise pour unité, quitte à rétablir ultérieurement l'homogénéité des formules, pour comparer avec les résultats du paragraphe D où cette distance était notée h). L'équation (22) donne alors, en remplaçant h_i et h_j par une valeur moyenne que nous noterons dorénavant h, et en faisant les mêmes approximations que celles qui avaient permis de passer de (8') à (8") :

$$(24) \begin{cases} 2h\varphi(0) - 4m[\varphi(1) - \varphi(0)] = \bar{\omega} \\ 2h\varphi(1) - 2m[\varphi(2) - \varphi(1)] - 2m[\varphi(o) - \varphi(1)] = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ 2h\varphi(n) - 2m[\varphi(n+1) - \varphi(n)] - 2m[\varphi(n-1) - \varphi(n)] = 0. \\ \dots\dots\dots\dots \end{cases}$$

en désignant par $\bar{\omega}$ la moyenne arithmétique, sur tous les groupes utilisés de $\omega_i(q_{i_m})$:

Ces équations linéaires sont -sauf la 1ère- sans second membre et fournissent donc pour $\varphi(n)$ une combinaison linéaire de k_1^n et de k_2^n, k_1 et k_2 étant les deux racines de "l'équation caractéristique", $2mk^2 - (4m + 2h)k + 2m = 0$ qui s'écrit encore : $k^2 - (2 + \frac{h}{m})k + 1 = 0$ et admet une seule racine inférieure à 1, que nous noterons k_1. Comme $\varphi(n)$ doit être borné, il doit être proportionnel à k_1^n, et cela même pour n = 0 d'après la 2ème équation (24). On a alors $\varphi(n) = \varphi(o)\, k_1^n$, et la 1ère équation (24) permet alors de calculer $\varphi(o)$ par la formule :

$$\varphi_o = \frac{\bar{\omega}}{2h + 4m(1 - k_1)}$$

Il sera commode, pour la simplicité des écritures, de supposer dorénavant que le taux h (qui caractérise grossièrement l'intensité de la mutation et de la sélection) soit petit % ou "taux de mutation" m [(1)]. On peut alors remplacer k_1 par $1 - \sqrt{\frac{h}{m}}$ et récrire $\varphi(n) = \varphi(o)\, k_1^n$ sous la forme :

----

(1) Cette hypothèse n'est pas indispensable ; les formules pourraient être écrites en s'en affranchissant, mais cela au prix de quelques longueurs. (cf. G. Malécot. Communication à la 31ème session de l'Institut International de Statistique, Bruxelles 1958).

(25) $$\varphi(n) = \frac{\bar{\omega}}{4\sqrt{mh}} \, 1^{-n\sqrt{\frac{h}{m}}}$$

$\bar{\omega}$ peut d'ailleurs être formulé, dans l'hypothèse où tous les groupes ont le même effectif Ni = N, la même corrélation gamétique $f_i = f$, et les mêmes variances V_u et V_v de leurs coefficients u_i et v_i et en désignant par M_1 et M_2 les moyennes arithmétiques, sur tous les groupes utilisés, des quantités $\bar{q}_{im}$ et $\bar{q}^2_{im}$ (quantités que l'on peut elles-mêmes assimiler respectivement à $\bar{q}_i$ et à $\bar{q}_i^2 = (\bar{q}_i)^2 + v_{ii}$) ; on a, alors :

(26) $$\bar{\omega} = \frac{(1 + f)(M_1 - M_2)}{2N} + V_u M_2 + V_v (1 - 2M_1 + M_2)$$

Comme M_2 dépend -d'ailleurs linéairement- de la moyenne arithmétique des v_{ii} qui est $\varphi(o)$, on voit que $\varphi(o)$ serait donné, comme au paragraphe D, par une équation linéaire ; nous ne la formerons pas ; puisque la fonction pratiquement observable est la fonction $\psi(x)$, il nous suffira de former, à la fin de notre étude, l'équation -de forme analogue- donnant $\psi(o)$. D'après les équations (23) et (25), la détermination de $\psi(x)$ est maintenant ramenée à celle de la fonction $\mu(x)$.

$\mu(x)$ étant la moyenne arithmétique des produits $(\bar{q}_i - \bar{q}_o)(\bar{q}_j - \bar{q}_o)$ correspondant à $j - i = x$, l'équation (21) (où l'on remplace h_i par la valeur moyenne h et l_{ki} par m quand $|k - i| = 1$ et par $1 - 2m$ quand $k - i = 0$) fournit :

$$\mu(x) = h \quad \text{moyenne des } [C_i - \bar{q}_o)(\bar{q}_j - \bar{q}_o)] + (1 - h)[(1 - 2m)\,\mu(x) +$$
$$+ m\,\mu(x - 1) + m\,\mu(x + 1)]$$

mais la résolution de cette récurrence nécessite la connaissance de la fonction de j-i = x qui est la moyenne arithmétique des $(C_i - \bar{q}_o)(\bar{q}_j - \bar{q}_o)$. Nous noterons $\theta(x)$ cette fonction et écrirons la récurrence précédente (en y négligeant le produit mh vis-à-vis de m) sous la forme :

(27) $$h\,\mu(x) - m\,[\mu(x + 1) - 2\,\mu(x) + \mu(x - 1)] = h\,\theta(x).$$

Cette récurrence avec 2ème membre ne peut être résolue que moyennant une détermination préalable de la fonction $\theta(x)$. Or celle-ci vérifie une récurrence sans 2ème membre que l'on obtient à partir de (21) [par soustraction de $\bar{q}_o$, multiplication par $(C_j - \bar{q}_o)$, et prise des moyennes arithmétiques] (1) :

---

(1) $\theta(x)$ est une fonction paire (invariante quand on échange i et j) du fait de la formule (27) et du fait que $\mu(x)$ est, d'après sa définition même, une fonction paire.

$$\theta(x) = h \quad \text{moyenne des } [(C_i - \bar{q}_o)(C_j - \bar{q}_o)] + (1 - h)[(1 - 2m)\theta(x) + m\,\theta(x-1) + m\,\theta(x+1)].$$

Si l'on admet que les variations des constantes $C_i = \frac{\overline{v_i}}{\overline{u_i} + \overline{v_i}}$ d'un groupe à l'autre ne sont pas influencées par la proximité géographique et ont donc dans l'espace géographique un caractère "purement aléatoire"(1) et si l'on particularise la "moyenne conventionnelle" $\bar{q}_o$ en la prenant égale à la moyenne arithmétique des C_i (ou, ce qui revient au même d'après (21), à la moyenne arithmétique des $\bar{q}_i$, que l'on peut estimer par la moyenne arithmétique des q'_i observés dans les différents groupes), la moyenne arithmétique des produits $(C_i - \bar{q}_o)(C_j - \bar{q}_o)$ n'est différente de zéro, et égale alors à la moyenne des $(C_i - \bar{q}_o)^2$ que nous désignerons par σ_c^2, que quand $j = i$, c'est-à-dire quand $x = 0$. On a alors :

(28) $$h\,\theta(x) - m[\theta(x+1) - 2\theta(x) + \theta(x-1)] = 0$$

sauf pour la valeur $x = 0$ qui donne $h\theta(o) - 2m[\theta(1) - \theta(o)] = h\sigma_c^2$. La résolution se fait par les mêmes formules que pour (24) où il suffit de remplacer $\bar{\omega}$ par $2h\sigma_c^2$. On obtient :

(29) $$\theta(x) = \frac{1}{2}\sqrt{\frac{h}{m}}\,\sigma_c^2\, e^{-|x|\sqrt{\frac{h}{m}}}$$

L'équation (27) peut alors être résolue ; mais du fait que son 2ème membre $h\theta(x)$ se trouve être une solution particulière de l'équation sans 2ème membre, on doit chercher la solution de (27) sous la forme :

(30) $$\mu(x) = [\alpha|x| + \beta]\, e^{-|x|\sqrt{\frac{h}{m}}}$$

On détermine le coefficient α par substitution dans (27), ce qui donne :

(30') $$\alpha = \frac{h}{4m}\,\sigma_c^2$$

---

(1) C'est l'hypothèse de "sélection par les micromilieux" de Sheppard et Cain. Dans le cas contraire où la proximité géographique créerait une corrélation entre les C_i, on pourrait introduire dans (28) des seconds membres proportionnels à cette corrélation, mais on doit, de toutes façons, exclure le cas où la variation des C_i présenterait un gradient géographique ("cline").

et le coefficient β n'est autre que $\mu(o)$ [qui est par définition de $\mu(x)$ moyenne arithmétique des $(\overline{q}_i - \overline{q}_o)^2$ et se calcule à l'aide de la formu (27) pour $x = 0$: $h\,\mu(o) - 2\,m\,[\mu(1) - \mu(o)] = h\,\theta(o)$; compte tenu de l'e pression (30) qui donne (pour $\frac{h}{m}$ petit) $\mu(1) - \mu(o) = \alpha - \sqrt{\frac{h}{m}}\,\beta$ on trouve :

(30")
$$\beta = \frac{1}{4}\sqrt{\frac{h}{m}}\,\sigma_c^2$$

On peut alors tirer des équations (23) et (25) l'expression de $\psi(x)$:

(31)
$$\psi(x) = \left[\frac{\overline{\omega} + h\,\sigma_c^2}{4\sqrt{mh}} + |x|\,\frac{h}{4\,m}\,\sigma_c^2\right] e^{-|x|\sqrt{\frac{h}{m}}}$$

Mais cette formule ne permet pas à elle seule de calculer $\psi(o) = \frac{\overline{\omega} + h\,\sigma_c^2}{4\sqrt{mh}}$ car $\overline{\omega}$ dépend de $\psi(o)$: en vertu de (26) où l'on remplace M_1 par $\overline{q}_o$ et M_2 par $\overline{q}_o^2 + \psi(o)$ (1) $\overline{\omega}$ s'écrit :

$$\frac{1+f}{2N}\,\overline{q}_o(1 - \overline{q}_o) + V_u\,\overline{q}_o^2 + V_v\,(1 - \overline{q}_o^2) - \left[\frac{1+f}{2N} - V_u - V_v\right]\psi(o).$$

$\psi(o)$ est donc fourni par une équation du 1er degré qui donne :

(32)
$$\psi(o) = \frac{\frac{1+f}{2N}\,\overline{q}_o(1 - \overline{q}_o) + V_u\,\overline{q}_o^2 + V_v\,(1 - \overline{q}_o)^2 + h\,\sigma_c^2}{\frac{1+f}{2N} + 4\sqrt{mh} - V_u - V_v}$$

et $\psi(x)$ admet alors l'expression [plus simple que (31)] :

(33)
$$\psi(x) = \left[\psi(o) + |x|\,\frac{h}{4\,m}\,\sigma_c^2\right] e^{-|x|\sqrt{\frac{h}{m}}}$$

Ces expressions englobent comme cas particulier celles qui avaient été obtenues dans les modèles précédemment considérés.

1/ Si l'on suppose que les taux u_i et v_i qui définissent les "forces de rappel linéaires" (2) sont constants dans le temps ($V_u = V_v = 0$) et constants d'un groupe à l'autre ($C_i = \frac{\overline{v}_i}{\overline{u}_i + \overline{v}_i}$ indépendant de i, donc

----

(1) $\overline{q}_o$ est la moyenne des quantités q_{i_m}, et leur variance est (à un terme de l'ordre de m près) égale à $\psi(o)$.

(2) Rappelons que ces taux tiennent compte à la fois de la mutation et (approximativement) de la sélection dans chaque groupe.

$\sigma_c^2 = 0$), on obtient la formule (33') $\psi(x) = \frac{\bar{q}_o (1 - \bar{q}_o)}{1 + 8N\sqrt{mh}} e^{-|x|\sqrt{\frac{h}{m}}}$ analogue à celles du paragraphe D, et qui se ramène, quand $\frac{h}{m}$ est petit, à celle que j'avais indiquée en 1949 [(1)] (Colloque de Lyon sur le Calcul des Probabilités, éditions du C. N. R. S.) et qui a été utilisée par Lamotte [(2)] pour déterminer, pour le gène, "absence de bandes" des Cepaea, le rapport $\frac{h}{m} = \frac{k}{\mu} = \frac{u + v}{\mu}$ entre le coefficient de rappel (k) dû aux pressions internes dans chaque groupe et le coefficient de migration (μ). En prenant pour unité d'évaluation de la distance x le kilomètre, qui est aussi la distance moyenne entre deux "colonies" voisines, Lamotte a obtenu pour $\frac{h}{m}$ une estimation de l'ordre de l'unité (ce qui traduit le fait que le coefficient de corrélation moyen $\frac{\psi(x)}{\psi(o)} = e^{-|x| \frac{h}{m}}$ entre deux colonies distantes de x décroît très rapidement d'une valeur voisine de 1 à une valeur voisine de zéro quand la distance x passe de 0, 1 km [(3)] à 2 km). Mais les observations de Cain et Sheppard, sur l'augmentation de variance que peut provoquer la variabilité des coefficients de sélection d'un groupe à l'autre, paraissent empêcher d'égaler la variance expérimentale à la variance théorique $\psi(o)$ qui résultent de (33'), et par suite d'estimer par cette voie le produit mh (comme l'avait fait Lamotte)(op. cit.).

2/ Si l'on fait intervenir, pour tenir compte de ces critiques, des taux u_i et v_i variant d'un groupe à l'autre ($\sigma_c^2 \# 0$), et si l'on suppose, pour simplifier, que ces taux restent constants au cours du temps dans chaque groupe ($V_u = V_v = 0$), la formule (33) permet, en ajustant l'expression théorique de $\frac{\psi(x)}{\psi(o)}$ aux coefficients de corrélation déterminés expérimentalement par Lamotte, une estimation de $\frac{h}{m}$ qui coïncide très sensiblement (la variation de l'exponentielle l'emportant beaucoup ,

----

(1) avec $f = o$, $h = u + v$, $q_o = C$: le 1er terme (l'unité) au dénominateur de (33'), y avait été négligé, ce qui n'est légitime que quand $\sqrt{mh}$ est grand % à $1/N$.

(2) Bulletin biologique de France et de Belgique, suppl. XXXV (1951) : h est noté k et m est noté μ.

(3) L'évaluation pour $x = 0, 1$ km résulte des données de Lamotte, qui avait observé par petits groupes de 10 m de côté, les individus d'une colonie très vaste : celle de Coquerel qui s'étend sur près d'un hm carré : il y a sensiblement panmixie sur toute son étendue.

dans la zone utilisée, sur celle de la fonction linéaire) avec celle faite au 1/ ; h est encore de l'ordre de grandeur de m.

Par contre, la formule (32) (avec $V_u = V_v = 0$ et $f = 0$) peut fournir si σ_c^2 est important, une valeur numérique (32'') : $\psi(o) = \frac{\bar{q}_o(1-\bar{q}_o) + 2Nh\sigma_c^2}{1 + 8N\sqrt{mh}}$ très supérieure à celle trouvée au 1/. Nous allons tenter de préciser l'influence possible de σ_c^2 (qui est la variance qu'aurait l'ensemble des valeurs d'équilibre locales en l'absence des migrations qui tendent à les égaliser) dans une région géographique telle que la portion d'Aquitaine particulièrement étudiée par Lamotte (1). Les colonies bien isolées y ont un effectif moyen de l'ordre de $N = 10^3$, et la variance de l'ensemble de leurs fréquences -variance qui est une estimation de $\psi(o)$- est de l'ordre de 10^{-2}. L'équation (32'') donne en y prenant $\bar{q}_o(1 - \bar{q}_o) = 0,2$ et $h = m$ (données de Lamotte, arrondies) :

$$\frac{1}{100}(1 + 8\,Nh) = 0,2 + 2\,Nh\sigma_c^2 \text{ d'où :}$$

(34) $$Nh = \frac{0,1}{0,04 - \sigma_c^2}$$

Si, avec Lamotte, on prend $\sigma_c^2 = 0$ et $N = 10^3$, on trouve $h\ (= u + v) = \frac{1}{400} = 0,0025$: c'est, en ordre de grandeur, ce qu'il indique p. 209 (2) ; mais, si σ_c^2 est $\neq 0$, l'estimation fournie pour h par la formule (34) peut devenir considérablement plus grande : il suffit pour cela que σ_c^2 soit voisin de 0,04, donc σ_c voisin de 0,2, ce qui se produirait si la variation d'une colonie à une autre du rapport $C_i = \frac{v_i}{u_i + v_i}$ pouvait atteindre l'ordre de $\frac{2}{10}$: une telle variation, combinée avec des coefficients de rappel u_i et v_i très élevés, (donc une sélection à la fois très importante et très variable) serait tout aussi compatible avec les résultats numériques actuellement disponibles, que la sélection faible et non variable admise par Lamotte. La discrimination entre les deux théories ne sera possible que si l'on parvient à une détermination directe du taux de migration m, qui serait aussi, comme nous l'avons vu, une estimation de h. Mais nous regardons comme vraisemblable la théorie de Lamotte, qui considère σ_c^2 comme négligeable (dans une région géographique suffisamment restreinte) car, dans le cas con-

----

(1) Op. cit. chap. X.

(2) Il nous paraît difficile d'isoler, comme il l'a fait, p. 207, la mutation de la sélection.

traire, m et h auraient des valeurs numériques nettement plus grandes que celles qu'il donne, alors que le taux de migration entre colonies, de quelques millièmes, qu'il a obtenu, est déjà fort élevé par rapport au déplacement quadratique moyen (de 15 mètres par animal) qu'il avait observé à l'intérieur d'une colonie homogène (1). Nous sommes donc enclin à conclure comme lui, au rôle secondaire du phénomène de sélection sur le gène "absence de bandes". Mais nous devons souligner à nouveau qu'une détermination directe de m apparaît comme indispensable pour trancher (2).

La formule (34) nous montre encore que l'estimation de h et m devrait être augmentée si celle de N était diminuée, c'est-à-dire si l'on tenait compte des fluctuations d'effectif dans certaines colonies instables (fluctuations qui se traduisent, comme l'a montré Sewall Wright et comme le rappelle Lamotte, par la nécessité de diminuer l'évaluation de l'effectif "efficace"). Le peu de vraisemblance d'une valeur trop élevée de h et m conduit encore à admettre que, dans les colonies d'Aquitaine étudiées par Lamotte, les nombres d'individus sont assez stables pour que l'effectif efficace ne soit pas très inférieur à 10^3. (Pour l'étude de "l'effectif efficace" et des problèmes que pose son évaluation, nous renvoyons à la suite du présent mémoire).

3/ Il est intéressant d'examiner aussi le cas particulier de (32) et (33) dans lequel σ_c^2 serait nul sans que V_u et V_v le soient, c'est-à-dire le cas de variations aléatoires de u_i et v_i au cours du temps dans chaque colonie -leurs moyennes étant toutefois les mêmes pour toutes les colonies. La loi de décroissance de la corrélation en fonction de la distance reste numériquement la même qu'au 1/, et l'estimation du rapport $\frac{h}{m}$ pour le gène "absence de bandes" chez les Cepaea d'Aquitaine, reste voisine de 1. Mais $\psi(o)$ est plus grand qu'au 1/, il est donné (si l'on suppose égales les variances V_u et V_v et si l'on fait $f = 0$) par :

----

(1) Remarquons incidemment que ceci nous oblige à admettre que les déplacements deviennent beaucoup plus importants pour les quelques individus qui ont, fortuitement, ou non, quitté une colonie, et que les probabilités, pour eux -ou pour leurs descendants- d'en atteindre une autre devient beaucoup plus grande.

(2) Indiquons à ce propos que la détermination, entreprise par le Dr Sutter, de taux de migration dans les populations humaines permettrait, conjointement avec une étude de la variabilité génétique, des déterminations indirectes de taux de mutation et de sélection chez l'homme, par des méthodes analogues à celles qui ont été employées pour les Cepaea.

$$\psi(o) = \frac{\bar{q}_o(1 - \bar{q}_o) + 2NV_u\ [\bar{q}_o^2 + (1 - \bar{q}_o)^2]}{1 + 8N\ \sqrt{mh} - 4\ NV_u}$$

En faisant $\psi(o) = 10^{-2}$ et $N = 10^3$ (avec $\bar{q}_o = 0,77$) on trouve :

$$\sqrt{mh} = 0,0025 + 15\ V_u$$

Si les coefficients u_i et v_i fluctuaient avec des écarts-types de l'ordre de 10^{-3}, V_u serait de l'ordre de 10^{-6} et son influence sur les valeurs numériques de m et h serait négligeable. Si les écarts-types de u_i et v_i étaient, par contre, de l'ordre de 10^{-2}, l'estimation à adopter pour m et h ne dépasserait pas le double de celle du 1/ et ne pourrait atteindre 10^{-2}, ce qui est contradictoire, car les moyennes de u_i et v_i, et par suite h, doivent être au moins de l'ordre de grandeur des écarts-types. La présente hypothèse n'a donc aucune valeur explicative en ce qui concerne les Cepaea (au paragraphe E, dans le "modèle de Kimura", nous avons déjà indiqué qu'il n'y a de modification importante que quand la variance du taux de sélection dépasse l'ordre de u ou celui de $\frac{1}{N}$, ce qui ne peut être le cas ici).

4/ Après ces quelques exemples, nous pouvons résumer le rôle des différents facteurs qui interviennent dans (32) et (33) en disant que le degré de différenciation à courte distance est caractérisé (1) par le rapport $\frac{h}{m}$ qui figure dans l'exponentielle : la différenciation entre groupes voisins est accrue par les pressions "internes" qui s'exercent dans chacun, et diminuée par la migration entre eux, alors que le degré de différenciation à grande distance est caractérisé par la variance $\psi(o)$ (2) qui est diminuée (pour $\frac{h}{m}$ fixé) par une augmentation de l'effectif N ou du taux de migration m, mais est par contre augmentée par une fluctuation (dans le temps, et surtout dans l'espace) des taux de mutation et de sélection (3). Nous avons ainsi pu compléter, par l'ana-

---

(1) Pour une valeur fixée de $j - i = x$, la moyenne arithmétique des différences $(q'_j - q'_i)^2 = [q'_j - \bar{q}_o - (q'_i - \bar{q}_o)]^2$ est $2(\psi(o) - \psi(x)$, dont le rapport a la variance $\psi(o)$ dépend essentiellement de $\frac{h}{m}$.

(2) Quand x est suffisamment grand, la moyenne des différences $(q'_j - q'_i)^2$ est égale à $2\ \psi(o)$.

(3) C'est pour cette raison que l'introduction de l'hypothèse de Sheppard, pour rendre compte de la variance observée par Lamotte, nécessiterait, pour compenser la fluctuation de ces taux, un coefficient de migration plus élevé encore que celui calculé par Lamotte.

lyse des incidences de la migration, les études fondamentales de R. A. Fisher et de Sewall Wright sur l'incidence conjointe de la mutation, de la sélection, et de l'effectif local (1).

Remarque finale :

Nous avons raisonné comme si les valeurs sélectives de gènes (A et a) occupant un même locus étaient indépendantes des gènes présents dans les autres loci. Cela n'est admissible que dans une région géographique sur l'étendue de laquelle les fluctuations génétiques ne sont pas trop importantes. Par contre, si l'on considère deux régions suffisamment séparées pour qu'il s'y soit accumulé des stocks génétiques très différents (par exemple l'Aquitaine et la Bretagne, en ce qui concerne les Cepaea), la valeur sélective d'un même gène peut être très différente d'une région à l'autre. Le système de gènes le mieux adapté à chaque région réalise un équilibre global, une "homeostasie" (Lerner), un "pic adaptatif" (S. Wright) qui peut différer d'une région à l'autre pour des raisons climatiques, écologiques, ou même fortuites. Mais nos connaissances insuffisantes sur les facteurs de l'adaptation rendent assez conjecturale une traduction mathématique précise (2).

G. MALECOT

---

(1) Indiquons encore que la formule (32) avec $o < f \leqslant 1$ entraîne la conséquence, souvent soulignée, que la pratique de croisements consanguins (en plus de ceux qui résulteraient de la panmixie) à l'intérieur du groupe local, équivaut à une réduction de son effectif dans un rapport qui ne peut être inférieur à $\frac{1}{2}$.

(2) Voir toutefois S. Wright (Journal of Genetics, 1935, p. 257).

CONCEPTION STOCHASTIQUE DE CŒFFICIENTS MULTIPLICATEURS DANS L'AJUSTEMENT LINÉAIRE DES SÉRIES TEMPORELLES

(MULTIPLICATIVE RANDOMNESS IN TIME-SERIES REGRESSION ANALYSIS)

H. THEIL et L. B. M. MENNES

Sommaire

I - INTRODUCTION -

Il est bien connu que l'analyse classique de l'ajustement linéaire, lorsqu'elle est appliquée à des séries temporelles économiques, crée un grand nombre de problèmes difficiles. L'un d'eux est la corrélation des résidus dans le temps ; un deuxième, le problème des saisons lorsque sont utilisées des unités de temps inférieures à une année ; un troisième, le caractère interdépendant et non contrôlé du développement de nos variables dans le temps, qui peut nous contraindre à adopter une approche par les équations simultanées ; un quatrième, l'effectif petit ou moyen de l'échantillon qui est à la base de nos calculs ; un cinquième, la présence d'erreurs d'observation ; un sixième, le fait gênant que le fondement économique de notre travail est d'ordinaire une micro-théorie, alors qu'en fait, nous avons besoin de macro-théories puisque nos variables sont habituellement des macro-variables. Or la difficulté principale n'est pas que cette liste de doléances puisse être allongée si facilement, mais le fait que la solution de l'un de ces problèmes est fréquemment entravée par les autres. Par exemple, si nous avons affaire à un système d'équations simultanées, nous pouvons avoir le désir d'augmenter l'effectif de notre échantillon, parce que l'analyse de l'estimation de tels systèmes est dans une large mesure une analyse asympto-

tique[1]. Mais si nous l'augmentons, disons en travaillant sur des observations trimestrielles au lieu de nos observations annuelles, nous trouvons en face de nous :

1/ - un problème de saisons,

2/ - une corrélation plus prononcée des perturbations successives, et

3/ - le fait que les organismes statistiques consacrent souvent moins d'efforts à la qualité des données trimestrielles qu'à celle des données annuelles.

Ce qui vient d'être dit montre que si l'on essaye d'attaquer isolément un problème particulier (par exemple le problème des équations simultanées dans l'hypothèse où il n'y a pas d'erreurs d'observation, où l'effectif de l'échantillon est suffisant, etc.), le résultat ne sera qu'une étape dans la direction visée. Cette remarque s'applique à la plupart des contributions récentes à la statistique économétrique ; et il n'est que juste de souligner qu'elle s'applique tout aussi bien à l'approche qui fait l'objet du présent article, à savoir le point de vue de la nature stochastique ou non stochastique des coefficients qui mesurent l'influence des variations des variables explicatives dans une analyse d'ajustement. Cette question peut être introduite à l'aide de l'exemple suivant :

$$y_t = \beta x_t + u_t \ , \qquad (1.1)$$

où y_t pour $t = 1, \ldots, T$ représente la variation[2] d'une variable dépendante quelconque pendant des périodes de temps successives, x_t la variation correspondante d'une variable explicative, β un coefficient mesurant l'effet de cette dernière variation sur la première, et u_t une perturbation aléatoire. Le problème ordinaire consiste alors à estimer β sur la base d'un nombre donné (T) d'observations sur les deux variables.

----

(1) Voir par exemple Theil [10, chapitre VI]. Récemment cependant, quelques essais couronnés de succès ont été effectués par A. L. Nagar [6, 7] sur l'estimation d'équations simultanées à partir d'échantillons finis.

(2) La plupart du temps, x et y sont définis comme des niveaux atteints par des variables, et non comme des variations de variables. Bien entendu, cette interprétation n'a aucune incidence sur le plan mathématique ; nous l'adoptons simplement parce qu'elle nous permet de formuler notre approche avec plus d'élégance.

Cette approche implique que les variations de la variable dépendante sont déterminées par deux forces distinctes. L'une de celles-ci est la perturbation aléatoire qui est supposée comprendre tous les facteurs qui affectent bien la variable dépendante mais ne sont pas introduits sous une forme explicite ; l'autre est la variable dont les variations sont indiquées par x, et qui est supposée influencer la variable dépendante en fonction d'un coefficient constant β. Or il est facile de voir que cette hypothèse de constance a de fortes chances d'être restrictive dans de nombreux cas. Supposons par exemple que x représente les variations du revenu réel par tête et y les variations de la consommation réelle par tête ; alors β (la propension marginale à consommer) mesure l'effet des variations du revenu sur les variations de la consommation dans l'hypothèse où cet effet est constant dans le temps. Mais on peut concevoir que cet effet ne soit pas constant et qu'il dépende de la question de savoir quels consommateurs voient leurs revenus affectés. Par exemple, si une augmentation du revenu par tête est imputable essentiellement à des accroissements des revenus des agriculteurs, nous devrions nous attendre à ce que l'effet soit moindre que dans le cas où il serait dû à des augmentations des revenus des travailleurs manuels dans les villes. Bien qu'il ne semble pas impossible de réduire de telles divergences par quelque désagrégation, le phénomène en tant que tel ne va pas disparaître.

Ces considérations conduisent à penser qu'il y a des circonstances où il est préférable de remplacer (1. 1) par la spécification plus générale :

$$y_t = b_t x_t + u_t \qquad (1.2)$$

où b_t est un coefficient aléatoire d'espérance β et de variance σ_1^2, ces derniers paramètres étant indépendants de t. Si $\sigma_1^2 = 0$, alors (1. 2) est équivalent à (1. 1) ; ce qui montre que (1. 2) est bien une spécification plus générale. En admettant que u_t soit lui aussi aléatoire (avec espérance zéro et variance σ_0^2), nous voyons que dans cette approche, il existe deux forces aléatoires affectant les variations de la variable dépendante.

Il y a quelques années, Hurwicz [3] et Rubin [8] ont envisagé le problème de l'estimation d'équations du type (1. 2), surtout pour signaler les difficultés qu'il comportait. Le présent article traite de l'estimation de l'espérance des coefficients aléatoires, et également de l'estimation des variances. Il concerne principalement le cas simple d'une seule variable explicative, étant donné que ce cas inclut toutes les particularités essentielles ; la généralisation est directe comme on le montrera dans la section 5. Des applications sont données dans la section 6.

II - HYPOTHESES -

Hypothèse 1 :

Pour chaque $t = 1, \ldots, T$, la variation y_t de la variable dépendante est donnée par (1. 2) dans le cadre des spécifications des hypothèses 2, 3, 4 et 5.

Les hypothèses 2 à 5 servent à éliminer l'influence perturbatrice d'autres complications (comme on l'a annoncé dans le second alinéa de la section 1). L'hypothèse 2 précise que les x sont fixés, ce qui exclut la possibilité qu'ils constituent une variable conjointement dépendante dans un système dont (1. 2) serait l'une des équations. Les hypothèses 3 et 4 précisent que nos variables aléatoires ont des distributions normales, cependant que l'hypothèse 5 exclut le cas où elles auraient une corrélation entre elles.

Hypothèse 2 :

Pour chaque $t = 1, \ldots, T$, la variation x_t de la variable indépendante est un nombre réel non stochastique. Les T valeurs $x_1, \ldots, x_T$ ne sont pas toutes égales.

Hypothèse 3 :

Pour chaque $t = 1, \ldots, T$, la perturbation additive u_t a une distribution normale de moyenne zéro et de variance σ_0^2 indépendante de t.

Hypothèse 4 :

Pour chaque $t = 1, \ldots, T$, le coefficient aléatoire b_t a une distribution normale d'espérance β et de variance σ_1^2 indépendante de t.

Hypothèse 5 :

Les 2T perturbations additives et coefficients aléatoires sont indépendants entre eux.

III - ESTIMATION DE L'ESPERANCE DES COEFFICIENTS ALEATOIRES -

Une méthode très simple pour estimer β est celle des moindres carrés :

$$b = \frac{\Sigma\, x_t\, y_t}{\Sigma\, x_t^2} \qquad (3.\ 1)$$

où Σ indique - ici et aussi bien plus loin - la sommation des T observa-

tions. Pour analyser plus commodément les propriétés de cet estimateur, nous allons écrire (1. 2) sous la forme :

$$y_t = \beta x_t + u_t^* \qquad (3.2)$$

où

$$u_t^* = u_t + (b_t - \beta) x_t \qquad (3.3)$$

En combinant les hypothèses 3, 4 et 5, nous voyons aisément que u_t^* a une distribution normale de moyenne zéro et de variance :

$$\text{Var } u_t^* = \sigma_0^2 + \sigma_1^2 x_t^2 \; ; \qquad (3.4)$$

et de plus que $u_1^*, u_2^*, \ldots, u_T^*$ sont indépendants entre eux.

En combinant (3. 1) et (3. 2), nous trouvons que b est un estimateur sans erreur systématique de β (1).

$$Eb = E \frac{\Sigma x_t (\beta x_t + u_t^*)}{\Sigma x_t^2} = \beta \qquad (3.5)$$

et que son erreur d'échantillonnage est :

$$b - \beta = \frac{\Sigma x_t u_t^*}{\Sigma x_t^2} \qquad (3.6)$$

En élevant au carré et en prenant la moyenne on a la variance :

$$\text{var } b = \frac{\Sigma x_t^2 \text{ var } u_t^*}{(\Sigma x_t^2)^2} = \sigma_0^2 \frac{1}{\Sigma x_t^2} + \sigma_1^2 \frac{\Sigma x_t^4}{(\Sigma x_t^2)^2} \qquad (3.7)$$

De même que dans le cas classique ($\sigma_1^2 = 0$), cette variance ne peut pas être déterminée exactement ; mais elle peut être estimée sans erreur systématique, parce que σ_0^2 et σ_1^2 peuvent elles-mêmes être estimées sans erreur systématique. On le montrera dans la section 4 ci-dessous.

La méthode des moindres carrés a l'avantage de la simplicité, mais elle n'est pas efficace. Le meilleur estimateur, linéaire sans erreur systématique est obtenu par les moindres carrés pondérés, méthode

----

(1) Ce résultat n'exige pas la condition que les coefficients aléatoires et les perturbations suivent une loi normale. Une condition suffisante est que les x soient des valeurs certaines, que les coefficients aléatoires aient tous la même espérance β, et que les perturbations additives aient toutes une espérance nulle.

où chaque observation est pondérée proportionnellement à l'inverse de la variance de u_t^*. Ceci conduit à :

$$b' = \sum \frac{x_t\, y_t}{\sigma_0^2 + \sigma_1^2\, x_t^2} \Big/ \sum \frac{x_t^2}{\sigma_0^2 + \sigma_1^2\, x_t^2} \qquad (3.8)$$

dont la variance est :

$$\text{var } b' = \left(\sum \frac{x_t^2}{\sigma_0^2 + \sigma_1^2\, x_t^2} \right)^{-1} \qquad (3.9)$$

ce qui est inférieur à la variance des moindres carrés (3. 7) sauf quand $\sigma_1^2 = 0$, auquel cas les deux méthodes d'estimation sont équivalentes. La méthode étudiée présente le désavantage évident que l'estimateur ponctuel ne peut être calculé que lorsque σ_0^2 et σ_1^2 sont connus. Le seul moyen (approximatif) de l'appliquer consiste par conséquent à calculer une estimation de β (par ex. le b des moindres carrés, à utiliser cette estimation dans le procédé qui sera exposé dans la section 4 pour obtenir des estimations de σ_0^2 et σ_1^2, et ensuite à appliquer (3. 8) avec σ_0^2 et σ_1^2 remplacés par leurs estimations.

IV - ESTIMATION DES VARIANCES -

Alors qu'il est possible dans le cadre de la spécification classique (1. 1) de trouver des estimations des perturbations u_t, ce n'est plus le cas dans le cadre de la spécification présente : ni les perturbations additives u_t de (1. 2), ni les déviations des coefficients aléatoires b_t par rapport à leur espérance β ne peuvent être estimées séparément. Ceci ne signifie pourtant pas qu'il soit impossible de trouver des estimations des paramètres de leurs distributions. Nous avons déjà trouvé des estimateurs sans erreur systématique de l'espérance des coefficients aléatoires, et il va s'avérer possible de trouver des estimateurs sans erreur systématique pour les variances d'une manière analogue.

En considérant la perturbation "combinée" u_t définie dans (3. 3), nous observons d'abord que cette perturbation peut être estimée à partir des données dont nous disposons (comme égale à y_t - b x_t) ; en second lieu, que la variance de cette perturbation est une fonction de la variation de la variable explicative qui contient les paramètres inconnus σ_0^2 et σ_1^2, cf. (3. 4). Ceci conduit à penser que nous devrions envisager le résidu des moindres carrés,

$$v_t = y_t - bx_t = (b_t - \beta)\, x_t - (b - \beta)\, x_t + u_t \qquad (4.1)$$

et prendre son carré comme estimateur de la variance de u_t^*. On trouve facilement que l'espérance de ce carré est :

$$Ev_t^2 = \sigma_0^2 \left\{1 - \frac{x_t^2}{\Sigma x^2}\right\} + \sigma_1^2 x_t^2 \left\{1 - 2\frac{x_t^2}{\Sigma x^2} + \frac{\Sigma x^4}{(\Sigma x^2)^2}\right\} \qquad (4.2)$$

Nous pouvons donc écrire :

$$v_t^2 = \sigma_0^2 P_t + \sigma_1^2 Q_t + w_t , \qquad (4.3)$$

où

$$P_t = 1 - \frac{x_t^2}{\Sigma x^2} \qquad (4.4)$$

$$Q_t = x_t^2 \left\{1 - 2\frac{x_t^2}{\Sigma x^2} + \frac{\Sigma x^4}{(\Sigma x^2)^2}\right\} , \qquad (4.5)$$

et w_t est une variable aléatoire de moyenne zéro pour tous les t :

$$Ew_t = 0 \qquad (4.6)$$

Nous appellerons (4. 3) l'ajustement du résidu correspondant à la spécification (1. 2). Ceci fait apparaître le carré de chaque résidu des moindres carrés comme fonction linéaire de deux fonctions connues des x, mis à part un terme additif (w_t) qui est la perturbation de l'ajustement du résidu. Etant donné que les x et par conséquent les P et les Q sont des valeurs certaines et que de plus les perturbations w_t ont pour moyenne zéro, l'application des moindres carrés à l'ajustement du résidu donne des estimations sans erreur systématique de σ_0^2 et σ_1^2, qui sont précisément les paramètres qui nous intéressent. Cependant, les moindres carrés ne sont pas nécessairement efficaces, puisque les w peuvent avoir une matrice de covariance qui diffère du type scalaire. En fait, nous avons :

$$\text{var } w_t = E(v_t^2 - Ev_t^2)^2 = Ev_t^4 - (Ev_t^2) = 2(Ev_t^2)^2 = 2(\sigma_0^2 P_t + \sigma_1^2 Q_t)^2 \qquad (4.7)$$

en tirant parti de l'existence d'une loi normale de v_t [v_t est une fonction linéaire de variances normales d'après (4. 1)], et du fait que le quatrième moment centré d'une variable normale est égal à trois fois le carré de la variance[1]. De la même façon, nous trouvons pour la covariance de w_t et $w_{t'}$ $(t \neq t')$:

$$\text{cov}(w_t, w_{t'}) = E[(v_t^2 - Ev_t^2)(v_{t'}^2 - Ev_{t'}^2)] =$$

$$= E(v_t^2 v_{t'}^2) - (Ev_t^2)(Ev_{t'}^2) = 2[E(v_t v_{t'})]^2 =$$

$$= 2\frac{x_t^2 x_{t'}^2}{(\Sigma x^2)^2}\left\{\sigma_0^2 + \sigma_1^2 x_t^2 + \sigma_1^2 x_{t'}^2 - \sigma_1^2 \frac{\Sigma x^4}{\Sigma x^2}\right\}^2 , \qquad (4.8)$$

----

(1) Voir par exemple Kendall [4, p. 80].

où on tire parti de ce que[1], si X et Y suivent une loi normale liée à moyennes nulles (comme v_t et $v_{t'}$), on a $E(X^2 Y^2) = EX^2\ EY^2 + 2\ [E(XY)]^2$. On voit immédiatement que la covariance est de l'ordre de $1/T^2$; et par conséquent, puisque les variances sont de l'ordre de 1 d'après (4. 7), que les perturbations de l'ajustement du résidu sont liées par une corrélation de l'ordre de $1/T^2$ seulement. Pour prendre un exemple, si nous considérons le cas où toutes les valeurs x sont égales[2], la corrélation d'un couple quelconque $(w_t, w_{t'})$ est égale à $1/(T-1)^2$, ce qui est une quantité tout-à-fait infime pour presque n'importe quel effectif d'échantillon réaliste. Nous allons donc négliger ces corrélations.

Le procédé d'estimation proposé pour σ_0^2 et σ_1^2 consiste donc à appliquer d'abord les moindres carrés ordinaires à (4. 3), ce qui conduit aux premières estimations s_0^2 et s_1^2 :

$$\begin{aligned} \Sigma v_t^2 P_t &= s_0^2\ \Sigma P_t^2 + s_1^2\ \Sigma P_t Q_t \\ \Sigma v_t^2 Q_t &= s_0^2\ \Sigma P_t Q_t + s_1^2\ \Sigma Q_t^2 \ ; \end{aligned} \qquad (4.\ 9)$$

puis on se sert de ces estimations pour appliquer les moindres carrés conformément à l'expression de la variance (4. 7) :

$$\begin{aligned} \sum \frac{v_t^2 P_t}{(s_0^2 P_t + s_1^2 Q_t)^2} &= s_0'^2 \sum \frac{P_t^2}{(s_0^2 P_t + s_1^2 Q_t)^2} + s_1'^2 \sum \frac{P_t Q_t}{(s_0^2 P_t + s_1^2 Q_t)^2} \\ \sum \frac{v_t^2 Q_t}{(s_0^2 P_t + s_1^2 Q_t)^2} &= s_0'^2 \sum \frac{P_t Q_t}{(s_0^2 P_t + s_1^2 Q_t)^2} + s_1'^2 \sum \frac{Q_t^2}{(s_0^2 P_t + s_1^2 Q_t)^2} \end{aligned} \qquad (4.\ 10)$$

$s_0'^2$ et $s_1'^2$ étant les moyennes estimations de σ_0^2 et de σ_1^2 respectivement. La matrice de covariance est :

$$V \begin{vmatrix} s_0'^2 \\ s_1'^2 \end{vmatrix} \approx 2 \begin{bmatrix} \sum \frac{P_t^2}{(s_0^2 P_t + s_1^2 Q_t)^2} & \sum \frac{P_t Q_t}{(s_0^2 P_t + s_1^2 Q_t)^2} \\ \sum \frac{P_t Q_t}{(s_0^2 P_t + s_1^2 Q_t)^2} & \sum \frac{Q_t^2}{(s_0^2 P_t + s_1^2 Q_t)^2} \end{bmatrix}^{-1} \qquad (4.\ 11)$$

----

(1) Voir par exemple Kendall [4, p. 80].

(2) Le cas spécial où les x sont égaux a été exclu dans l'hypothèse 2, aussi cet exemple n'a-t-il qu'une valeur illustrative. On voit facilement que ce cas doit être rejeté, car sinon P_t et Q_t seraient constants (c'est-à-dire indépendants de t), si bien qu'il serait impossible d'estimer σ_0^2 et σ_1^2.

Le signe ≈ est employé pour indiquer que la matrice de covariance n'est égale qu'approximativement à la matrice de droite. L'égalité serait exacte si s_0^2 et s_1^2 étaient remplacées par leurs valeurs vraies, dans (4. 10) et dans (4. 11).

Il y a lieu de faire trois remarques supplémentaires. D'abord, bien que les moindres carrés généralisés appliqués à (4. 3) donnent les meilleures estimations linéaires(1) sans erreur systématique de σ_0^2 et σ_1^2 si on veut apprécier leur signification, on est gêné par le fait que l'emploi traditionnel des écarts-types tels qu'on les obtient à partir de (4. 11) n'est pas parfaitement satisfaisant. La raison en est que la variable dépendante de la régression du résidu est non-négative, en sorte qu'elle ne peut pas être distribuée suivant une loi normale, et la même remarque s'applique aux estimations de σ_0^2 et σ_1^2 ; par conséquent, il n'est pas entièrement correct de dire que l'estimation ponctuelle $s_0'^2$ plus ou moins deux fois (par exemple) l'écart-type, contient σ_0^2 avec une probabilité obtenue à partir d'une loi de Student. Le deuxième point est que l'ajustement du résidu se caractérise par une dispersion considérable, ce qui implique qu'un assez grand nombre d'observations est en général nécessaire pour estimer avec précision σ_0^2 et σ_1^2. Pour bien voir ceci, considérons l'équation $v_t^2 = Ev_t^2 + w_t$; après avoir élevé au carré à gauche et à droite, pris l'espérance mathématique, et pris la moyenne dans le temps, nous obtenons :

$$\frac{1}{T}\, \Sigma_t\, Ev_t^4 = \frac{1}{T}\, \Sigma_t (Ev_t^2)^2 + \frac{1}{T}\, \Sigma_t\, Ew_t^2 . \qquad (4.\ 12)$$

En comparant les deux termes de droite et en appliquant (4. 7), nous trouvons que le second terme est égal au double du premier. Par conséquent $R^2 = \frac{1}{2}$ pour l'ajustement du résidu, R étant la corrélation multiple vraie de cet ajustement(2). On peut se servir de ceci comme test de la suffisance des hypothèses de base ; mais on n'entrera pas ici dans cette voie. Le troisième point enfin est étroitement lié au second : l'écart-type de la perturbation de l'ajustement du résidu est proportionnel à l'espérance de sa variate de gauche, le facteur de proportionnalité étant $\sqrt{2}$. Il est intéressant de noter que ce type d'hétéros-

----

(1) linéaires par rapport aux carrés des résidus v_t^2.

(2) On notera que R est fondé sur des moments ordinaires, non pas sur des moments centrés. Par conséquent $1 - R^2 (=2/3)$ est le rapport du second moment des perturbations de l'ajustement des résidus [= le deuxième terme dans le membre de droite de (4. 12)] au second moment des carrés des résidus [= le terme situé à gauche dans (4. 12)]. Le second moment des perturbations est évidemment identique à la variance, mais celui des v_t^2 ne l'est pas.

cédasticité s'applique aussi à certaines données sur les ménages,(1) l'écart-type de l'erreur aléatoire par rapport à la courbe d'Engel étant ici proportionnel à l'ordonnée correspondante de la courbe. Dans ce cas, la constante de proportionnalité n'a pas nécessairement la valeur $\sqrt{2}$ bien entendu.

V - SIMPLIFICATIONS ET GENERALISATIONS -

On notera que la différence entre P_t et 1 et celle entre Q_t et x_t^2 sont l'une et l'autre de l'ordre de 1/T [cf. (4. 4) et (4. 5)]. En conséquence, si nous sommes disposés à commettre une erreur de cet ordre de grandeur, nous pouvons simplifier (4. 3) pour avoir :

$$v_t^2 = \sigma_0^2 + \sigma_1^2 x_t^2 + w_t' \tag{5. 1}$$

où w_t' a une espérance de l'ordre de 1/T. Comme la différence entre w_t et w_t' est de l'ordre de 1/T, et comme en outre w_t a une variance de l'ordre de 1, on voit aisément que la variance de w_t' à la même valeur compte non tenu des termes d'ordre de grandeur en T inférieur. Nous avons donc le droit d'écrire :

$$\text{var } w_t' \approx 2\,(\sigma_0^2 + \sigma_1^2 x_t^2)^2 \,, \tag{5. 2}$$

parce qu'au niveau d'approximation où nous nous trouvons, $P_t \approx 1$ et $Q_t \approx x_t^2$. De même on peut montrer que les w_t' sont liés par une corrélation de l'ordre de $1/T^2$ seulement.

Cette simplification, qui suppose essentiellement que nous considérions v_t comme une approximation asymptotique de u_t^*, peut être permise quand l'effectif de l'échantillon est véritablement suffisant . Elle revêt une importance toute spéciale quand le nombre de variables explicatives n'est plus 1, comme nous l'admettons jusqu'à présent, mais davantage, disons 2. Dans ce cas, nous devons remplacer (1. 2) par :

$$y_t = b_{1t}\, x_{1t} + b_{2t}\, x_{2t} + u_t \tag{5. 3}$$

où x_{1t} et x_{2t} sont les variations de deux variables explicatives différentes. Nous étendrons à ce cas les hypothèses de la section 2 de façon que l'une et l'autre variables explicatives prennent des valeurs certaines, que u_t, b_{1t}, b_{2t} suivent une loi normale avec les moyennes 0, β_1, β_2 et les variances σ_0^2, σ_1^2, σ_2^2 respectivement, et que les 3T coefficients aléatoires et perturbations additives soient indépendants entre eux. Nous pouvons alors écrire (5. 3) sous la forme équivalente :

----

(1) cf. Theil [9] et Fisher [1].

$$y_t - \beta_1 x_{1t} + \beta_2 x_{2t} = u_t^* \qquad (5.4)$$

où la perturbation combinée :

$$u_t^* = u_t + (b_{1t} - \beta_1)\, x_{1t} + (b_{2t} - \beta_2)\, x_{2t} \qquad (5.5)$$

a une espérance nulle et la variance :

$$\text{var } u_t^* = \sigma_0^2 + \sigma_1^2\, x_{1t}^2 + \sigma_2^2\, x_{2t}^2 \,. \qquad (5.6)$$

Après avoir estimé β_1 et β_2 par les moindres carrés, nous obtenons les résidus des moindres carrés v_t. En appliquant l'argument de l'alinéa précédent, nous pouvons alors établir l'ajustement du résidu :

$$v_t^2 = \sigma_0^2 + \sigma_1^2\, x_{1t}^2 + \sigma_2^2\, x_{2t}^2 + w_t' \qquad (5.7)$$

où w_t' est une perturbation dont l'espérance est de l'ordre de $1/T$ et la variance :

$$\text{var } w_t' \approx 2(\sigma_0^2 + \sigma_1^2\, x_{1t}^2 + \sigma_2^2\, x_{2t}^2)^2 \,, \qquad (5.8)$$

si on néglige les termes d'ordre élevé. La méthode à suivre exposée pour le cas d'une seule variable est donc facilement généralisée, et l'extension au cas de trois variables ou plus est directe.

Il est également assez facile de généraliser les hypothèses 3 à 5 de façon que les T couples (u_t, b_t) deviennent des tirages au hasard indépendants effectués dans une population-mère normale à deux dimensions de moyennes $(0, \beta)$, de variances (σ_0^2, σ_1^2) et de covariance $\sigma_{01} \neq 0$. Ce cas offre la possibilité d'une corrélation du décalage additif et du décalage multiplicatif, soit positive, soit négative. Il nous faut alors remplacer (4. 2) par :

$$Ev_t^2 = \sigma_0^2 \left\{ 1 - \frac{x_t^2}{\Sigma x^2} \right\} + 2\sigma_{01}\; x_t \left\{ 1 - 2\frac{x_t^2}{\Sigma x^2} + x_t \frac{\Sigma x^3}{(\Sigma x^2)^2} \right\} +$$

$$+ \sigma_1^2 x_t^2 \left\{ 1 - 2\frac{x_t^2}{\Sigma x^2} + \frac{\Sigma x^4}{(\Sigma x^2)^2} \right\}$$

ce qui peut être simplifié pour donner :

$$Ev_t^2 \approx \sigma_0^2 + 2\sigma_{01}\; x_t + \sigma_1^2 x_t^2 \qquad (5.10)$$

si nous sommes disposés à négliger les termes de l'ordre de $1/T$. De même si nous avons deux variables explicatives (cf. l'alinéa précédent),

et si les T ensembles de trois variates (u_t, b_{1t}, b_{2t}) sont des tirages indépendants au hasard effectués dans une population-mère normale à trois dimensions de moyennes (0, β_1, β_2) et avec une matrice de covariance :

$$\begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 \end{bmatrix} \tag{5.11}$$

alors l'espérance du carré du résidu des moindres carrés est de l'ordre de $1/T$:

$$Ev_t^2 \approx \sigma_0^2 + \sigma_1^2 x_{1t}^2 + \sigma_2^2 x_{2t}^2 + 2(\sigma_{01} x_{1t} + \sigma_{02} x_{2t} + \sigma_{12} x_{1t} x_{2t}) \tag{5.12}$$

ce qui est une forme quadritique qui s'obtient avec (5. 11) comme matrice et ($1 \quad x_{1t} \quad x_{2t}$) comme vecteur.

VI - APPLICATIONS -

Nous avons appliqué le procédé proposé dans les sections 3 et 4 aux données sur les prix d'importation et d'exportation du Royaume-Uni pendant la période 1870-1952. La théorie (simple, on l'admet) est que la variation annuelle en pourcentage de l'indice des prix d'exportation (y) dépend de la variation en pourcentage de l'indice des prix d'importation (x) pour la même année selon (1. 2). L'effectif de l'échantillon est T = 65, les années de guerre étant exclues(1).

L'estimation de β par les moindres carrés (la première étape) est :

$$b = 0,795 \tag{6.1}$$

Les moindres carrés appliqués à l'ajustement du résidu (deuxième étape) (2) donnent :

$$s_0^2 = 11,5 \qquad \text{et} \qquad s_1^2 = 0,044 \tag{6.2}$$

----

(1) Les données de base, qui ont été prises dans Kindleberger [5, pp. 22-25] concernent les importations et les exportations de marchandises ; elles ont été reproduites dans l'Appendice (Tableau 1).

(2) Le carré de la corrélation multiple correspondant à cette spécification de la régression du résidu est $R^2 = 0,371$. En appliquant la valeur vraie de $R^2 = 1/3$ de notre hypothèse nulle, il n'y a aucune possibilité de rejet. Ainsi l'écart-type tel qu'il est donné par Kendall [4, p. 385, eq. (15. 64)] a la valeur 0,095. On notera, bien entendu, qu'une telle appréciation de la signification est contrariée à la fois par l'absence de loi normale et par l'hétéroscédastivité.

En appliquant les moindres carrés pondérés à l'ajustement du résidu (troisième étape), nous trouvons :

$$s'^2_0 = 9,7 \qquad \text{et} \qquad s'^2_1 = 0,075 \tag{6.3}$$

les écarts-types étant alors d'après (4.11) respectivement 2,4 et 0,033. En utilisant les critères traditionnels, nous dirons que les deux valeurs (6.3) sont positives dans une mesure significative. Enfin, l'estimateur de β par les moindres carrés pondérés fondé sur les estimations ponctuelles (6.3) est :

$$b' = 0,855 \tag{6.4}$$

avec un écart-type de 0,078. L'écart-type de la première estimation b, également fondée sur (6.3) est 0,098. En nous en tenant aux estimations ponctuelles, nous pouvons conclure que la variation en pourcentage du niveau des prix d'exportation est soumise à une perturbation additive qui suit une loi de moyenne zéro et d'écart-type 3.1, et qu'elle est affectée par la variation en pourcentage du niveau des prix d'importation en fonction d'un coefficient aléatoire d'espérance 0.85 et d'écart-type 0,27.

On a déjà souligné le fait qu'un assez grand nombre d'observations est en général nécessaire pour réaliser avec succès cette approche. Ceci s'applique bien entendu en particulier lorsqu'il y a plusieurs variables explicatives. L'exemple suivant est illustratif à cet égard ; il concerne la consommation réelle par tête aux Etats-Unis, qui est exprimée linéairement en fonction du revenu réel par tête courant et du revenu réel par tête le plus élevé atteint dans le passé(1). La relation a pour base des données annuelles pour la période 1897-1949, les années de guerre exclues ; l'effectif de l'échantillon est T = 44. En appliquant les moindres carrés à (5.3) (x_1 et x_2 étant interprétés respectivement comme les variations du revenu courant et du revenu le plus élevé atteint dans le passé, nous obtenons :

$$b_1 = 0,531 \qquad \text{et} \qquad b_2 = 0,228 \tag{6.5}$$

---

(1) Le revenu a été défini comme le revenu personnel disponible, d'après les données de Goldsmith [2, Vol. III, p. 429] ; la consommation est obtenue en soustrayant l'épargne personnelle totale [2, Vol. I, p. 370] l'une et l'autre séries ont été divisées par la population totale résidant aux Etats-Unis d'après les Statistical Abstracts. Les deux séries ainsi obtenues sont reproduites dans l'Appendice (tableau 2).

L'ajustement du résidu est alors appliqué sous sa forme simplifiée (5. 7) et les estimations ponctuelles par les moindres carrés sont

$$s_0^2 = 482 \qquad s_1^2 = -0{,}065 \qquad s_2^2 = 0{,}119 \qquad (6.6)$$

s_0 étant mesuré en dollars de 1929. Il est clair que la valeur négative de s_1^2 est mauvaise et que la valeur positive de s_2^2 est d'une grandeur peu vraisemblable. Cet exemple montre aussi qu'il vaut la peine de commencer par mettre en œuvre l'ajustement du résidu sous sa forme simplifiée, car la forme complète nécessite beaucoup plus de travail de la part du calculateur.

REFERENCES BIBLIOGRAPHIQUES

(1) - FISHER, G. R. - "Maximum Likelihood Estimators with Heteroscedastic Errors". Review of the International Statistical Instituto. Vol. 25 (1957), pp. 52-55.

(2) - GOLDSMITH, R. W. - A Study of Saving in the United States. Three Volumes (Princeton, 1955-1956).

(3) - HURWICZ, L. - "Systems with Nonadditive Disturbances". Chapter VIII of Statistical Inference in Dynamic Economic Models, edited by T. C. Koopmans (New-York - London, 1950).

(4) - KENDALL, M. G. - The Advanced Theory of Statistics. Vol. I, 3rd edition (London, 1947).

(5) - KINDLEBERGER, C. P. - The Terms of Trade. A European Case Study (New York, 1956).

(6) - NAGAR, A. L. - "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations". Report 5 820 of the Econometric Institute of the Netherlands School of Economics (1958).

(7) - NAGAR, A. L. - "Simultaneous Equations Estimation : The Bias of the Residual Variance Estimator of a Single Equation". Report 5 823 of the Econometric Institute of the Netherlands School of Economics (1958).

(8) - RUBIN, H. - "Note on Random Coefficients". Chapter XIX of Statistical Inference in Dynamic Economic Models, edited by T. C. Koopmans (New-York - London, 1950).

(9) - THEIL, H. - "Estimates and their Sampling Variance of the Parameters of certain Heteroscedastic Distributions". Review of the International Statistical Institute, Vol. 19 (1951), pp. 141-147.

(10) - THEIL, H. - Economic Forecasts and Policy (Amsterdam, 1958).

APPENDICE

Tableau 1 : Indices de la valeur unitaire des importations (P_M) et des exportations (P_X) de marchandises pour le Royaume-Uni 1870-1952 (1913 = 100)

Années	P_M	P_X	Années	P_M	P_X
1870	120	104	1905	86	84
1871	115	104	1906	90	89
1872	123	117	1907	94	94
1873	123	121	1908	90	91
1874	120	113	1909	92	88
1875	116	106	1910	99	92
1876	113	98	1911	97	94
1877	113	93	1912	98	96
1878	106	89	1913	100	100
1879	102	85			
1880	107	88	1920	214	270
1881	106	85	1921	151	213
1882	105	85	1922	138	182
1883	103	84	1923	140	180
1884	98	81	1924	141	173
1885	92	77	1925	154	183
1886	86	74	1926	142	173
1887	86	73	1927	136	165
1888	88	75	1928	137	162
1889	89	78	1929	134	159
1890	88	82	1930	117	151
1891	89	82	1931	88	126
1892	86	78	1932	64	91
1893	85	77	1933	74	110
1894	78	74	1934	93	132
1895	75	72	1935	93	130
1896	77	73	1936	98	135
1897	77	72	1937	111	145
1898	77	72	1938	103	147
1899	79	77			
1900	85	89	1948	245	296
1901	82	85	1949	229	283
1902	82	81	1950	194	221
1903	84	82	1951	258	260
1904	85	83	1952	252	273

Tableau 2 : Revenu réel disponible par tête (Y) et consommation (C) pour les Etats-Unis 1897-1949 (en dollars de 1929).

Années	Y	C	Années	Y	C
1897	367.2	350.0	1923	594.1	507.7
1898	369.3	329.5	1924	596.5	521.7
1899	388.0	327.0	1925	595.0	504.0
1900	396.7	363.2	1926	622.4	539.0
1901	430.7	395.6	1927	629.9	545.3
1902	426.6	353.7	1928	619.3	569.9
1903	436.9	401.1	1929	670.5	576.2
1904	434.8	401.1	1930	616.5	569.0
1905	449.9	370.6	1931	580.9	558.1
1906	495.5	427.8	1932	491.3	524.9
1907	506.6	465.7	1933	478.8	519.2
1908	454.3	413.4	1934	513.9	523.5
1909	524.9	465.7	1935	561.6	538.8
1910	499.5	440.0	1936	631.7	578.9
1911	488.8	451.6	1937	644.9	579.9
1912	488.4	418.9	1938	600.0	565.2
1913	496.1	453.2	1939	643.2	579.9
1914	463.1	423.0	1940	678.7	601.6
1915	496.9	425.2	1941	783.3	668.0
1916	527.6	451.8			
			1946	879.5	752.9
1919	602.0	517.5	1947	823.6	722.7
1920	539.4	490.1	1948	835.0	713.5
1921	500.1	488.8	1949	822.9	721.9
1922	527.9	470.1			

FONCTIONS DE RÉPARTITION A N DIMENSIONS ET LEURS MARGES

M. SKLAR

M. M. Fréchet [1] a étudié le problème de la détermination des fonctions de répartition à n dimensions dont les marges sont données. En ce qui suit nous considérons ce problème d'un point de vue nouveau.

Pour les définitions des fonctions de répartition, des marges, etc., voir p. 67 dans le premier ouvrage cité en (1). Les définitions qu'on y trouve sont exprimées pour n = 2, mais elles peuvent être étendues à n quelconque.

THEOREME 1 -

Soit G_n une fonction de répartition à n dimensions avec les marges F_1, F_2, ..., F_n. Soit R_k l'ensemble des valeurs de F_k, pour k = 1, 2, ..., n. Alors il existe une fonction unique H_n, dont l'ensemble de définition est le produit Cartésien $R_1 \times R_2 \times \ldots \times R_n$ et telle que :

$$G_n(x_1, \ldots, x_n) = H_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

<u>Définition 1</u> :

Nous appellerons <u>copule</u> (à n dimensions) toute fonction C_n continue et non-décroissante [2] définie sur le produit Cartésien de n intervalles fermés [0, 1] et satisfaisant aux conditions :

---

(1) M. Fréchet : sur les tableaux de corrélation dont les marges sont données, Ann. Univ. Lyon, Sect. A (3), 14, 1951, 53-77. Voir aussi M. Fréchet, Comptes Rendus Acad. Sci., Paris, 242, 1956, 2426-2428; E. J. Gumbel, Ibid., 246, 1958, 2712-2719 ; R. Féron, Publ. Inst. Statistique Univ. Paris, 5, 1956, 3-12.

(2) "Non-décroissante" au sens employé pour une fonction de répartition à n dimensions.

$$C_n(0, 0, \ldots, 0) = 0, \; C_n(1, \ldots, 1, \alpha, 1, \ldots, 1) = \alpha$$ [1].

THEOREME 2 -

La fonction H_n de Théorème 1 peut être prolongée (en général de plusieurs façons) à une copule C_n. Etant un prolongement de H_n, la copule C_n satisfait à la condition :

$$G_n(x_1, \ldots, x_n) = C_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$$

THEOREME 3 -

Soient données des fonctions de répartition $F_1, F_2, \ldots, F_n$ à une dimension. Soit C_n une copule quelconque à n dimensions. Alors la fonction G_n définie par :

$$G_n(x_1, \ldots, x_n) = C_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$$

est une fonction de répartition à n dimensions avec les marges $F_1, F_2, \ldots, F_n$.

Les Théorèmes 1-3 réduisent le problème mentionné de M. Fréchet au problème de la caractérisation des couples à n dimensions. Il n'y a qu'une seule copule à une dimension, notamment P_1, qui satisfait la condition $P_1(x) = x$ pour tout x entre 0 et 1. Pour $n > 1$, il y a une infinité de copules de dimension n, limitées par deux fonctions particulières.

THEOREME 4 -

Une copule quelconque C_n à n dimensions satisfait aux inégalités :

$$L_n \leqslant C_n \leqslant M_n,$$

où :

$$L_n(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathrm{Max}(\alpha_1 + \alpha_2 + \ldots + \alpha_n - n + 1, 0)$$
$$M_n(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathrm{Min}_n(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Pour $n \geqslant 1$, les fonctions P_n, définies par :

$$P_n(\alpha_1, \alpha_2, \ldots, \alpha_n) = \alpha_1 \alpha_2 \ldots \alpha_n \quad (0 \leqslant \alpha_k \leqslant 1, \; k = 1, 2, \ldots, n)$$

sont des copules à n dimensions. Elles déterminent le tableau de corrélation de n variables aléatoires indépendantes.

---

(1) Des cas spéciaux de telles fonctions ont été considérés par M. Féron, op. cit.

Or, les copules sont en général d'une structure plus simple que les fonctions de répartition. On peut utiliser cette simplicité en introduisant la notion du "quasi-inverse" d'une fonction monotone.

Soit f une fonction monotone d'une variable réelle. Ajoutons à l'ensemble des points $(x, f(x))$ tous les segments fermés de la forme $[(x, f(x+)), (x, f(x-))]$. Réfléchissons l'ensemble résultant dans la ligne $y = x$. Enlevons de l'ensemble réfléchi tous les segments verticaux à l'exception d'un seul point sur chacun d'eux. Tout ensemble qu'on peut obtenir de cette façon représente une fonction, que nous nommerons une quasi-inverse de la fonction f [1].

Selon cette définition, chaque fonction de répartition F à une dimension possède au moins un quasi-inverse F^*, défini et non-décroissant sur l'intervalle fermé $[0, 1]$ et prenant des valeurs finies sur l'intervalle ouvert $(0,1)$. Parmi ces quasi-inverses, il y en a un qui est continu à gauche sur $(0, 1)$ et continu à 0 et à 1.

THEOREME 5 -

Soit Q_n une fonction de n variables, intégrable dans l'espace euclidien E^n à n dimensions par rapport à une fonction de répartition G_n à n dimensions. Soit C_n une copule telle que l'on ait :

$$G_n(x_1, x_2, \ldots, x_n) = C_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$$

où $F_1, F_2, \ldots, F_n$ sont les marges de G_n. Alors :

(1) $$\int_{E^n}\!\!\ldots\!\int Q_n(x_1, x_2, \ldots, x_n)\, d\, G_n(x_1, x_2, \ldots x_n)$$

$$= \int_{U^n}\!\!\ldots\!\int Q_n(F_1^*(\alpha_1), F_2^*(\alpha_2), \ldots, F_n^*(\alpha_n))\, d\, C_n(\alpha_1, \alpha_2, \ldots, \alpha_n),$$

où U^n est le produit Cartésien de n intervalles $[0, 1]$.

L'expression au côté droit de la formule (1) peut être plus simple que celle au côté gauche. C'est le cas, par exemple, si C_n est absolument continue, comme au cas $C_n = P_n$; ou $n = 2$ et $C_2 = L_2$ ou M_2 ; ou si $n = 3$ et $C_3 = M_3$.

----

(1) Des formes diverses de telles fonctions quasi-inverses ont été considérées par plusieurs auteurs.

UN MODÈLE ÉCONOMÉTRIQUE

LA PROJECTION A LONG TERME DE VERDOORN

P. THIONET

1/- M. Verdoorn a dit dans un article assez récent d'Econométrica (1) comment le Bureau Central de Planification de La Haye s'y était pris pour effectuer une projection à long terme du revenu national des Pays-Bas ; cette projection s'étend de 1950 à 1970 ; mais les derniers résultats des calculs n'ont été publiés qu'en 1955 de sorte qu'il s'agit plutôt d'une projection à 15 ans qu'à 20 ans de distance. La méthode de Verdoorn semble avoir été décrite dans un travail en hollandais publié dès 1952. Le "modèle de Verdoorn" a fait l'objet en 1958-59 de discussions dans des réunions internationales.

On discutera ici exclusivement l'article paru dans Econometrica, lequel doit représenter l'aspect le plus élaboré de conceptions susceptibles d'une évolution notable. Il ne s'agit manifestement pas d'une œuvre mûrie dans l'abstrait mais d'un travail qui s'est imposé par les besoins de la planification et qui ne se soucie guère d'esthétique mathématique.

2/- Son point de départ présente de sérieuses ressemblances avec les conceptions "globales" de nombreux économistes (français ou non) : on se donne comme but de déterminer le revenu national global en 1970.

Là-dessus se greffe bien une opération "Input-Output" ; c'est-à-dire que, l'économie étant supposée ventilée (disons) en onze branches (2) on calcule les taux d'expansion (1970/1950) auxquels chacune de ces

----

(1) P.J. Verdoorn - Complementarity and longrange projections Econometrica Oct. 1956, p. 429 à 450.

(2) Métaux, Chimie, Textiles et habillement, Alimentation, Services Publics, Mines, Bâtiment et Travaux Publics, Autres industries, Agriculture, Transports, Autres services.

branches arrive ; mais ce n'est pas du tout le Tableau d'échanges industriels qui a permis de dire que ces taux sont atteints en 1970, plutôt disons qu'en 1980.

3/- Nous allons donc étudier un modèle très global ; mais ce n'est pas en fait un vice : les Pays-Bas ne sont pas du tout en mesure de vivre en quasi autarcie (1).

Dès lors il est secondaire de se donner ou non des objectifs de consommation de chaque type de produits en 1970. L'essentiel est de produire des articles exportables ; et ce n'est justement pas là qu'on peut raffiner beaucoup le modèle.

Bref la planification ne demande pas du tout à être traitée par les mêmes moyens aux Pays-Bas, en France ou en U.R.S.S.

4/- Entrons plus avant dans l'étude du modèle (2). Pour Tinbergen, les équations du 1er degré étaient susceptibles de remplacer en première approximation à peu près toute autre équation, mais ce qui est bon pour des problèmes de statique ne vaut plus lorsqu'on étudie des évolutions portant sur 20 ans.

Ainsi Verdoorn a conservé partout les fonctions puissances et les exponentielles sans les "linéariser". Mais bien des relations (les identités comptables surtout) restent des relations linéaires ; ces deux types de relations combinés ne peuvent rien donner de simple.

En particulier et comme une sorte de seconde approximation, les courbes d'équation mal connue sont remplacées partout par des courbes à élasticité constante. Or la plupart du temps, on devra ajouter ou retrancher entre elles ces fonctions à exposants disparates.

---

(1) Voici l'importance relative de la production et des importations dans les Pays-Bas et en France.

(Vers 1952-1955)	Production	Importations non agricoles
	(Milliards de dollars)	
Pays-Bas	3,1	1,55
France	12,4	1,8

(2) Celui-ci est reproduit en annexe.

Il n'est donc pas surprenant que le modèle de Verdoorn, basé sur 17 relations très simples (beaucoup trop simples, même) entre 19 variables aboutisse à des équations résolvantes (après élimination des variables intermédiaires) d'aspect peu engageant.

Analysons à présent les 3 parties du modèle : Capital, Travail, Balance extérieure.

5/- Le Capital : Sa loi de formation.

La partie du modèle concernant le Capital est assez subtile.

- une fraction déterminée du revenu national constitue l'épargne (dont une partie ad libitum peut être exportée suivant les besoins en devises ; le reste est investi).

- une fraction constante $(1 - \omega)$ de la dépréciation de l'équipement existant (20 % dans l'application numérique du modèle) vient s'ajouter à l'équipement neuf (qu'on appelle équipement "net") et représente le matériel amorti mais encore capable de servir.

Mais la seconde de ces fractions serait une vraie constante, alors que la première serait une variable instrument entre les mains des pouvoirs publics (on l'intitule à cet effet : propension à épargner).

Pour mieux comprendre le modèle du Capital nous allons imaginer que le pays vit en économie fermée et regarder ce que cela implique.

Le modèle comporte les deux concepts de Produit National Brut (PNB) v et de Revenu National y, qui diffèrent par une quantité d appelée la Dépréciation (1) :

$$y = v - d$$

On écrit donc :

Investissements nets : $i = \alpha y$

Investissements de Remplacement : $i' = \omega d$; $d - i' = (1 - \omega) d$

Si k est le stock de capital et $\dot{k}$ la dérivée de k par rapport au temps, on écrit :

---

(1) "Depreciation" en anglais.

Formation d'équipements :

$$\dot{k} = i + d - i'$$
$$= \alpha(v - d) + d(1 - \omega)$$

ou : $$\boxed{\dot{k} - \alpha v = (1 - \omega - \alpha)\, d}$$

Ici il faut marquer un temps d'arrêt.

Avec le modèle Keynesien "naïf", où la dépréciation et les remplacements sont supposés identiques l'un à l'autre, on écrit couramment :

$$\dot{k} - \alpha v = 0$$

Verdoorn remplace au 2ème membre zéro par un terme non identiquement nul, mais qui en fait est petit.

La valeur numérique du paramètre ω finalement retenue par Verdoorn est 0,80, donc $1 - \omega = 0{,}20$. Ainsi $(1 - \omega - \alpha)$ est assez petit, car la propension à épargner α est assez voisine de 20 %.

Par rapport au modèle naïf, le modèle de Verdoorn apporte un certain freinage à la constitution d'équipements lorsque α devient trop fort supérieur à $(1 - \omega)$ et au contraire pousse à accroitre le stock d'équipements lorsque α est trop faible, inférieur à $(1 - \omega)$.

6/- La demande d'équipements.

Tout modèle oppose la demande à l'offre ; et ce qu'on vient d'écrire est l'équation de l'offre. La demande d'équipements est imprimée par l'expansion du P.N.B. ; le phénomène physique est celui des goulots d'étranglement ; mais c'est la psychose de l'entrepreneur qui détermine les commandes d'équipements neufs, et il n'est pas absurde d'assimiler celui-ci à un consommateur d'équipements, et de supposer une certaine permanence à l'élasticité de la demande d'équipement ; soit χ cette élasticité qu'on choisit d'ailleurs ici égale à 80 %.

Si le P.N.B. est en train d'augmenter disons de 5 % l'an, le stock d'équipements tendra à augmenter de 4 % l'an.

(Ce qui surprend, c'est que la production puisse toujours physiquement augmenter ainsi de 5 %).

Il nous manque une équation : celle qui fixe l'importance des dépréciations. Là encore on admet qu'une loi psychologique nous condamne à une élasticité constante, et on prend cette fois l'élasticité (ν) égale à 120 % (par rapport au stock de capital).

On observera que l'élasticité des dépréciations (d) par rapport au P.N.B. est de ce fait :

$$1.20 \times 0,8 = 0.96 \text{ ou } 96\ \%$$

$$\text{pratiquement}\ldots\ldots\ 100\ \%$$

7/- L'équation différentielle.

Il est clair que cet ensemble de relations implique des conséquences mathématiques. En éliminant les d et k, il reste une équation différentielle en v de la forme :

$$\boxed{\dot{v} = A\, v^{n} + B\, v^{n'}}, \qquad \text{(V)}$$

On trouve effectivement :

$$d = d_o^* k^{\nu} \qquad k = K v \qquad \begin{cases} A = \alpha / Kx \\ B = (1 - \alpha - \omega)\, d_o^* K^{\nu}/Kx \\ n = 2 - x \\ n' = x^{\nu} - x + 1 \end{cases}$$

Cette équation n'est pas en général élémentaire. Il est facile d'en faire une intégration approchée.

Pour cela remarquons qu'on aurait $n = n'$ pour $x\nu = 1$; ce qui serait le cas avec :

$$x = 0,8 \text{ et } \nu = 1,25$$

Il nous a paru intéressant de savoir ce qui se passerait si l'on faisait :

$$\nu = 1,25 \text{ au lieu de } \nu = 1,20$$

dans le modèle de Verdoorn, sans rien changer au reste. Il vient :

$$\dot{v} = (A + B)\, v^{n} \qquad \begin{cases} \text{avec } n = 2 - 0.80 = 1.20 \\ \text{donc } 1 - n < 0 \end{cases}$$

d'où :

$$v_o^{1-n} - v^{1-n} = (A + B)\ (n - 1)\, t$$

Le P.N.B. devient infini au bout d'un temps fini T, donné par l'équation :

$$v_o^{1-n} = (A + B)\,(n - 1)\, T$$

Afin de nous assurer que T est très supérieur à 20 ou 30 ans, faisons (1) (avec Verdoorn) :

$$v_o = 1 ;\ n - 1 = 0,2 ;\ K = 5 ;\ K^{\nu} = 7,477.$$

Essayons : $\alpha = 20\ \%$; il vient :

$$A = 0,05 ;\ B = 0 ;\ A + B = 0,05$$

d'où : $$T = (0.05 \times 0.2)^{-1} = \boxed{100 \text{ ans}}$$

Essayons : $\alpha = 8\ \%$; il vient :

$$A = 0,0200 ;\ B = 0,0029 ;\ A + B = 0,0229$$

d'où : $$T = (0,023 \times 0,2)^{-1} = \boxed{220 \text{ ans}} \text{ environ}$$

Concluons que le modèle ne serait plus valable à l'échelle séculaire.

L'allure des courbes trouvées est caractérisée par leur asymptote verticale.

Le produit national brut a une croissance plus rapide que la fonction exponentielle ; son taux de croissance annuel n'est pas fixe mais croissant.

<u>Recherche des solutions à taux de croissance fixe.</u>

L'équation (V) a une solution exponentielle si (et seulement si), en posant $v = \exp(wt)$:

a) Les exposants de $\dot{v}$, v^n et $v^{n'}$ coïncident, d'où $w = wn = wn'$

où : $$w = w\ (2 - x) = w\ [1 + x\ (\nu - 1)]$$

où : $$\boxed{\nu = x = 1}$$

b) En outre les coefficients des deux membres sont égaux, ce qui fait (avec $x = 1$) :

$$w = A + B$$

En prenant $d_o^* = 1, w = \frac{\alpha}{K} + 1 - \alpha - \omega$, avec $K = 5$, (1)... il vient :

----

(1) Valeurs numériques empruntées à Verdoorn lui-même.

$$\boxed{v = \exp\,(1 - \omega - 0.8\,\alpha)\,t}$$

Nous nous étonnerons que le taux d'expansion de l'économie puisse être $(1 - \omega - 0,8\,\alpha)$, c'est-à-dire être d'autant plus faible que la propension à épargner est grande.

Rappelons le modèle naïf (avec $1 - \omega - \alpha = 0$)

$$\dot{k} - \alpha v = 0$$

pour lequel on aurait bien : $v = v_o \exp wt$; $k = K \exp wt$; donc :

$$w = \frac{v_o\,\alpha}{K}$$

avec : $K = 5$; $v_o = 1$; $\alpha = 20\ \%$; $w = 4\ \%$.

Lorsqu'on n'a plus :

$$\boxed{0 = 1 - \omega - \alpha}$$

le terme $\lambda = 1 - \omega - \alpha$ n'est pas négligeable à côté de α/K.

Par ex.

$\alpha = 18\ \%$; $\lambda = 2\ \%$; $\frac{\alpha}{K} = 3,6\ \%$; $w = 5,6\ \%$

$\alpha = 22\ \%$; $\lambda = 2\ \%$; $\frac{\alpha}{K} = 4,4\ \%$; $w = 2,4\ \%$

$\alpha = 25\ \%$; $\lambda = 2\ \%$; $\frac{\alpha}{K} = 5,0\ \%$; $w = 0\ \%$

L'économie serait stagnante $\alpha = 25\%$ (parce que l'excès d'épargne entrainerait une dépréciation trop forte). A l'opposé, pour $\alpha = 0$, on aurait $w = 1 - \omega = 20\%$ (en l'absence de toute épargne).

On voit à quels paradoxes on aboutit.

8/- Observations importantes.

Le Modèle de Verdoorn comprend (on l'a dit) 3 éléments : Capital, Travail, Devises. On vient de voir comment à chaque valeur de α (propension à épargner) le premier fait correspondre une équation différentielle définissant la loi de croissance du produit national brut ; on a même une idée de la forme des courbes représentant v en fonction de

t en économie fermée : On a vu qu'elles n'étaient pas exponentielles, ayant une asymptote verticale et croissant plus vite qu'une exponentielle. Il est donc exclu que l'économie ait un taux de croissance uniforme (même à la limite au bout d'un temps assez long). Ceci parait d'ailleurs naturel ; car nous savons combien sont irréalistes les modèles économiques (du genre de celui de Von Neumann) où chaque secteur a le même taux d'expansion que l'ensemble (lequel taux est également le taux d'intérêt. Les taux d'expansion de chaque secteur diffèrent entre eux et varieront dans le temps ; leur agrégat n'a nul besoin d'une croissance uniforme.

Cependant on va, avec la partie du modèle concernant le Travail, trouver de nouvelles conditions liant le temps, le Produit National Brut (v) et le Revenu National (y), ceci pour un taux d'émigration donné.

En fin de compte, les pouvoirs publics sont censés agir sur 2 variables instruments : (α) propension à épargner et (φ) taux d'émigration ; et s'il n'était pas procédé constamment à l'ajustement de l'un des instruments sur l'autre, il n'y aurait aucune raison pour que la valeur de (v) (calculée plus haut) soit identique (à tout instant) à la valeur de (v) qui va résulter des ressources en main d'oeuvre. La solution de l'équation différentielle donnée plus haut ne serait donc l'image des faits qu'autant qu'on ferait émigrer chaque année juste assez de hollandais pour maintenir constante la propension à épargner. (1).

9/- Le Modèle et le Travail.

L'offre de travail : (à population constante).

On suppose que l'offre de travail (en heures) et le revenu moyen par tête (revenu national divisé par l'effectif de la population) sont liés par une loi à élasticité constante ψ : élasticité prise égale à - 0,25.

Cette offre est par ailleurs proportionnelle à la population N, dont les mouvements de longue durée sont symbolisés par une variation exponentielle du temps.:

- taux d'accroissement naturel π (1,2 % par an) ;
- moins : taux d'émigration φ (variable-instrument).

La demande de travail.

On suppose que la demande (bien entendu égale à l'offre) est liée

---

(1) Il est facile d'expliciter v (t) pour φ constant si on le désire.

au P.N.B. (v) par une autre loi à élasticité constante (0,275 = ρ).

Remarques.

1/- Le travail et le capital, nos deux facteurs de production, sont supposés rigoureusement complémentaires. Aucune substitution du capital au travail n'est prévue. Ainsi l'expansion économique dans un pays à population constante serait bloquée dès le plein emploi atteint.

2/- Verdoorn admet (avec réticence) que dans certains cas, on doive introduire dans le modèle le phénomène : accroissement de la productivité, c'est-à-dire : progrès technique. Il conçoit celui-ci sous forme d'une exponentielle du temps corrigeant la loi de Demande de Travail. On manque d'applications numériques.

10/- Le Modèle et la Balance extérieure : Les variables.

Les 19 variables (15 endogènes et 4 instruments) liées par 17 équations, laissent deux degrés de liberté. Séparée du monde extérieur, l'économie aurait un degré de liberté, représenté par l'instrument α (propension à épargner) - et par l'instrument φ (taux d'émigration nette), entre qui existe une relation (du moment que l'économie est en équilibre).

Quand on réintroduit le Monde Extérieur dans le modèle, le gouvernement dispose de 2 autres instruments, qui sont désignés par γ et σ :

γ symbolise la politique d'exportations, c'est le taux de variation à long terme du rapport p/p'' des prix à l'exportation réels et des prix à l'exportation "compétitifs".

σ symbolise au contraire la politique d'importations, c'est le taux de variation à long terme de la demande d'importations.

En pratique Verdoorn a étudié des valeurs de σ nulles ou négatives (- 4 %) et au contraire des valeurs de γ de part et d'autre de 0, pour rechercher la politique optimale (voir § 12) et a d'ailleurs trouvé comme optimum pratiquement $\gamma = 0$.

A propos de cette même balance extérieure interviennent encore 3 variables exogènes (la quatrième étant l'accroissement démographique naturel de 1.2 % par an, mentionné plus haut).

a) Taux d'accroissement de la demande autonome importée : $\beta = 0{,}039$.

Pour les Pays-Bas, (vu leur faible superficie et leurs ressources

minérales réduites) l'expansion économique implique un développement permanent des importations.

b) Indice des prix à l'importation. $p'' = 1$ (à toute époque) :

Il s'agit manifestement d'une convention commode.

c) Taux de variation du ratio p''/p' : $\delta = -0.0025$

Les prix compétitifs à l'exportation p" sont rapportés aux prix à l'importation p'. Il est admis qu'à long terme le prix des produits industriels (exportés) tend à baisser par rapport au prix des matières premières (importées).

11/- Le Modèle et la Balance extérieure : Les équations.

Sur les 17 équations du modèle, nous en comptons 7 dans cette partie. La demande d'importations est supposée obéir à une loi à élasticités partielles constantes : par rapport aux exportations et par rapport au P.N.B. Les exportations sont (de leur côté) élastiques par rapport au prix relatif (p/p") ; mais on leur a appliqué cette fois une formule d'élasticité "à la Tinbergen", où :

$$1 + \varepsilon\left(\frac{p}{p''} - 1\right) \text{ remplace } \left(\frac{p}{p''}\right)^{\varepsilon} \quad (\text{avec } \varepsilon = -2)$$

D'ailleurs les indices des prix des exportations ordinaires p et compétitives p" ne seront jamais dans un rapport très supérieur à 1, ce qui justifie l'approximation faite.

Dans cette partie du modèle on ne compte pas moins de 5 fonctions exponentielles du temps. Des 5 taux : 2 sont des instruments et 2 des variables exogènes (déjà citées) ; le 5ème taux λ est le taux de décroissance du rapport q, valeur des importations sur valeur des exportations ; il est choisi de manière à ramener en 20 ans ce rapport q de la valeur initiale 1,07 à une valeur tenue pour raisonnable (et qui n'est pas précisée) ; λ est ce qui est appelé une variable-objectif ("target").

Lorsqu'on considère une économie réelle (non isolée du monde) cette partie du modèle conduit [pour le P.N.B. v] à une expression explicite du temps, combinaison d'exponentielles assez simple, que nous avons facilement calculée numériquement.

$$v = \left(3 - 2\,e^{\gamma t}\right)^{0.83} \cdot e^{t\,8/0.89}$$

L'instrument γ symbolise la politique d'exportations à long terme.

Le terme $B = 0.0264 + \gamma - \sigma + \lambda$ représente la résultante de 2 instruments γ et σ, d'un objectif λ, et d'élasticités structurelles dont les valeurs supposées ont permis de calculer 0.0264 ; γ et σ sont négatives ou positifs, λ est négatif ; tous 3 sont nécessairement petits. Ainsi nous n'osons trop affirmer que B sera encore positif ; à coup sur il sera très petit.

a) Cas de γ négatif :

Des 2 facteurs composant v, le premier est nécessairement croissant : à supposer B nul, v tendra vers la limite $3^{0.83} = 2,5$ (sachant v égal à 1) lorsque le temps t deviendra infini.

Si B est positif, v sera croissant et tendra (exponentiellement) vers l'infini.

Si B est négatif, v cessera un jour de croitre et prendra le chemin de tendre vers zéro ; tel serait le résultat d'une politique d'austérité trop rigoureuse, due aux difficultés de la balance en devises.

De toute façon, même si B est positif, ces fonctions v (t) croissent beaucoup moins vite que celles trouvées pour le modèle en économie fermée (§ 7).

b) Cas de γ nul :

Si B était négatif, v serait constamment et irrémédiablement décroissant. Il est plus raisonnable d'admettre B positif et v croissant (exponentielle à faible exposant).

c) Cas de γ positif :

Pour la même raison on admettra que B est positif ; le P.N.B. v croit d'abord mais finira par décroitre (et devrait même devenir négatif par la suite si le modèle avait encore un sens).

Là encore le contraste avec l'allure des courbes du § 7 est frappant. Ceci ne signifie pas que l'autarcie soit recommandable pour les Pays-Bas ; car il ne serait pas réaliste de supposer à ce pays des ressources naturelles illimitées de toutes les matières premières.

On arrive plutôt à concevoir la vraie trajectoire de v (t) comme une suite d'arcs de courbe empruntés aux 2 systèmes de courbes : un arc très rapidement croissant correspondant à un déficit en devises alternant avec un arc stagnant correspondant à un effort du côté de la balance des comptes, vu qu'il est difficile d'agir simultanément et correctement sur ces 2 sortes de leviers de commande.

12/- Le modèle et La notion de croissance optimale.

Donnant à la variable-instrument σ diverses valeurs, Verdoorn a procédé (pour une gamme de valeurs de γ) au calcul en 1970 du Revenu National par tête ; le graphique des résultats montre qu'il existe une valeur de γ qui maximise ce Revenu. On peut considérer que chaque couple $\sigma(\gamma)$ ainsi obtenu représente une politique optimale (1).

Ce calcul numérique a été fait pour divers jeux d'hypothèses sur les exportaticns autonomes en 1970 ; c'est-à-dire qu'au lieu de se cantonner à la variable exogène $\beta = 0.039$ (voir 10,a) on a essayé une gamme de valeurs de β entre 0.032 et 0.046.

CONCLUSION

L'énorme mérite du modèle de Verdoorn, c'est de mettre en lumière la totalité des hypothèses que comporte la projection ; il est honnête d'expliciter tout, jusqu'aux formules mathématiques inclusivement, non seulement pour autrui mais bien pour soi-même. Malheureusement le résultat n'est pas très élégant ni très clair.

Puisse la présente note avoir aidé à faire connaitre le modèle de Verdoorn, les modèles économiques et les problèmes de projection à long terme.

----

(1) Cet "optimum" coïncide heureusement dans le présent modèle avec le minimum d'émigration de hollandais. C'est certainement l'aspect le plus délicat du modèle.

ANNEXE

MODÈLE DE VERDOORN

Equations linéaires	(1) $v = c + i + d + b - m$	m : importations
	(2) $y = c + i + bp - m$	b : exportations
	(5) $i' = \omega d$	c : consommations
	(7) $i = \alpha y - b'$	
	(12) $m = qb\, p/p'$	
	(13) $b' = (1 - q)\, b p$	
Equation différentielle	(4) $k + i' = i + d$	
Equations à élasticité constante	(3) $k = K v^{x}$	
	(6) $d = d_o^* k^{\nu}$	
	(8) $a = v^{\rho}$	
	(9.1) $a_1 = a_o^* (y/N)^{\psi}$	parmi les variables, a_1 b_1 m_1 ne sont pas comptés
	(11.1) $m_1 = m_o b^{\eta} v^{\mu - \eta}$	
idem linéarisée	(15.1) $b_1 = b_o \left\{ 1 + \omega\left(\frac{p}{p''} - 1\right)\right\}$	
Croissances exponentielles	(9.2) $a = a_1 \exp(\pi - \varphi)t$	(9.1) et (9.2)
	(10) $N = \exp(\pi - \varphi)t$	(11.1) et (11.2)
	(11.2) $m = m_1 \exp \sigma t$	(15.1) et (15.2)
	(14) $q = q_o \exp \lambda t$	comptent pour 3 équations seulement
	(15.2) $b = b_1 \exp \beta t$	
	(16) $p = p'' \exp \gamma t$	
	(17) $p'' = p' \exp \delta t$	

Résultantes de Verdoorn

(I) $$y = v - d - b\,(1 - p)$$

(Capital) (II) $$\frac{v}{v} = \frac{1}{xk}\left\{\alpha y + (1 - \omega)\, d_o^* k^{\nu} - b'\right\}$$

(Travail) (III) $$v = v_o\, y^{\psi/\rho} \exp\; t(\pi - \varphi)\,(1 - \psi)/\rho$$

(Balance extérieure) (IV) $$v^{\mu-\eta} = [1 + \varepsilon\,(e^{\gamma t} - 1)]^{(1-\eta)}\; \exp t\,[\beta\,(1 - \eta) + \gamma - \sigma + \delta + \lambda]$$

(avec $a_o^* = v_o^{\rho}$)

CONTRIBUTION A LA THÉORIE DES VALEURS EXTRÊMES

Jean GEFFROY

ERRATUM

Page 49 THEOREME 5 -

La démonstration de ce théorème est erronée ; nous avons substitué in extremis cette fausse démonstration, par un souci abusif de symétrie avec le théorème 4, à la démonstration originale, qui est la suivante :

- En vertu du théorème 3, il nous suffit d'établir que si $Y_n^\alpha \underset{p}{\ll} b_n$, on a : $n\,G(b_n) \longrightarrow 0$. Si cette condition n'est pas satisfaite, il existe une suite infinie (n_i) de valeurs de n et un nombre $k > 0$ tels que :

$$n_i\,G(b_{n_i}) \longrightarrow k \text{ quand } i \longrightarrow \infty. \qquad (C)$$

Il est impossible que $k = +\infty$, car il en résulterait, d'après les théorèmes (3) et (4) :

$$b_{n_i} \underset{p}{\ll} Y_{n_i}^\alpha$$

En supposant k fini, la condition (C) entraine, d'après la formule (1) :

$$F_{y_{n_i}^\alpha}(b_{n_i}) \longrightarrow e^{-k}\left[1 + \frac{k}{1!} + \dots + \frac{k^{\alpha-1}}{(\alpha-1)!}\right] < 1$$

ce qui est en contradiction avec l'hypothèse : $Y_n^\alpha \underset{p}{\ll} b_n$.

Nous remercions bien vivement M. Alain Cohen d'avoir attiré notre attention sur cette erreur.

IMP. LOUIS-JEAN — GAP
Depôt légal n° 19 — 1960